PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/29, 15/82, 5/10, A01H 5/00

A2

(11) International Publication Number: WO 96/40924

(43) International Publication Date: 19 December 1996 (19.12.96)

(21) International Application Number: PCT/US96/09897

(22) International Filing Date: 7 June 1996 (07.06.96)

(30) Priority Data:
08/480,178 7 June 1995 (07.06.95) US

(60) Parent Application or Grant
(63) Related by Continuation
US Not furnished (CIP)
Filed on Not furnished

(71) Applicant *(for all designated States except US)*: CALGENE, INC. [US/US]; 1920 Fifth Street, Davis, CA 95616 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: McBRIDE, Kevin [US/US]; 1309 Marina Circle, Davis, CA 95616 (US). STALKER, David, M. [US/US]; 2736 Cumberland Place, Davis, CA 95616 (US). PEAR, Julie, R. [US/US]; 818 Douglass Avenue, Davis, CA 95616 (US). PEREZ-GRAU, Luis [ES/US]; 1230 Elk Place, Davis, CA 95616 (US).

(74) Agents: SCHWEDLER, Carl, J. et al.; Calgene, Inc., 1920 Fifth Street, Davis, CA 95616 (US).

(81) Designated States: AU, CA, CN, JP, KG, KZ, MX, TJ, TM, TR, US, UZ, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: COTTON FIBER TRANSCRIPTIONAL FACTORS

(57) Abstract

Novel DNA constructs are provided which may be used as molecular probes or inserted into a plant host to provide for modification of transcription of a DNA sequence of interest during various stages of cotton fiber development. The DNA constructs comprise a cotton fiber transcriptional initiation regulatory region associated with a gene which is expressed in cotton fiber. Also provided is novel cotton having a cotton fiber which has a natural color introduced by the expression in the cotton fiber cell, using such a construct, of pigment synthesis genes. Cotton fiber cells having color produced by genetic engineering and cotton cells comprising melanin and indigo pigments are included.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

COTTON FIBER TRANSCRIPTIONAL FACTORS

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation in part of United States application Serial No. 08/487,087 filed June 7, 1995, and a continuation in part of United States application Serial No. 08/480,178, filed June 7, 1995.

INTRODUCTION

Technical Field

This invention relates to methods of using *in vitro* constructed DNA transcription or expression cassettes capable of directing fiber-tissue transcription of a DNA sequence of interest in plants to produce fiber cells having an altered phenotype, and to methods of providing for or modifying various characteristics of cotton fiber. The invention is exemplified by methods of using cotton fiber promoters for altering the phenotype of cotton fiber, and cotton fibers produced by the method.

Background

In general, genetic engineering techniques have been directed to modifying the phenotype of individual prokaryotic and eukaryotic cells, especially in culture. Plant cells have proven more intransigent than other eukaryotic cells, due not only to a lack of suitable vector systems but also as a result of the different goals involved. For many applications, it is desirable

to be able to control gene expression at a particular stage in the growth of a plant or in a particular plant part. For this purpose, regulatory sequences are required which afford the desired initiation of transcription in the appropriate cell types and/or at the appropriate time in the plant's development without having serious detrimental effects on plant development and productivity. It is therefore of interest to be able to isolate sequences which can be used to provide the desired regulation of transcription in a plant cell during the growing cycle of the host plant.

One aspect of this interest is the ability to change the phenotype of particular cell types, such as differentiated epidermal cells that originate in fiber tissue, *i.e.* cotton fiber cells, so as to provide for altered or improved aspects of the mature cell type. Cotton is a plant of great commercial significance. In addition to the use of cotton fiber in the production of textiles, other uses of cotton include food preparation with cotton seed oil and animal feed derived from cotton seed husks.

Despite the importance of cotton as a crop, the breeding and genetic engineering of cotton fiber phenotypes has taken place at a relatively slow rate because of the absence of reliable promoters for use in selectively effecting changes in the phenotype of the fiber. In order to effect the desired phenotypic changes, transcription initiation regions capable of initiating transcription in fiber cells during development are desired. Thus, an important goal of cotton bioengineering research is the

acquisition of a reliable promoter which would permit expression of a protein selectively in cotton fiber to affect such qualities as fiber strength, length, color and dyability.

Relevant Literature

Cotton fiber-specific promoters are discussed in PCT publications WO 94/12014 and WO 95/08914, and John and Crow, Proc. Natl. Acad. Sci. USA, 89:5769-5773, 1992. cDNA clones that are preferentially expressed in cotton fiber have been isolated. One of the clones isolated corresponds to mRNA and protein that are highest during the late primary cell wall and early secondary cell wall synthesis stages. John and Crow, *supra*.

In animals, the *ras* superfamily is subdivided into the subfamilies *ras* which is involved in controlling cell growth and division, *rab/YPT* members which control secretory processes, and *rho* which is involved in control of cytoskeletal organization (Bourne et al., (1991) Nature 349: 117-127), and number of homologous genes have now been identified in plants (for a review, see Terryn et al., (1993) Plant Mol. Biol. 22: 143-152). None have been found for the important *ras* subfamily, all but one of the genes identified have been members of the *rab/YPT1* subfamily, and there is only one recent report of the cloning of a *rho* gene in pea (Yang and Watson(1993) Proc. Natl. Acad. Sci. USA 90: 8732-8736).

Little work has been done to characterize the functions of these genes in plants, although one recent report has shown that a small G protein from Arabidopsis can functionally complement a

mutant form in yeast involved in vesicle trafficking, suggesting a similar function for the plant gene (Bednarek et al., (1994) Plant Physiol 104: 591-596).

In animals, two members of the *rho* subfamily, called Rac and Rho, have been shown to be involved in the regulation of actin organization (for a review, see Downward, (1992) Nature 359: 273-274).

Rac1 has been shown to mediate growth factor-induced membrane ruffling by influencing microfilament alignment on the plasma membrane (Ridley et al, (1992) Cell 70: 401-410), whereas RhoA regulates the formation of actin stress fibers associated with focal adhesions (Ridley and Hall, (1992) Cell 70: 389-399).

In yeast, the CDC42 gene codes for a *rho*-type protein which also regulates actin organization involved in the establishment of cell polarity required for the localized deposition of chitin in the bud scar (Adams et al., (1990) J Cell Biol 111: 131-143.

Disruption of gene function, either by temperature shifts with a CDC42-temperature-sensitive mutant in yeast (Adams et al., 1990), or by micro-injection into fibroblasts of mutant Rac or Rho proteins exibiting a dominant negative phenotype (Ridley et al., 1992; Ridley and Hall, 1992), leads to disorganization of the actin network.

In plants, control of cytoskeletal organization is poorly understood in spite of its importance for the regulation of patterns of cell division, expansion, and subsequent deposition of secondary cell wall polymers. The cotton fiber represents an excellent system for studying cytoskeletal organization. Cotton

fibers are single cells in which cell elongation and secondary wall deposition can be studied as distinct events. These fibers develop synchronously within the boll following anthesis, and each fiber cell elongates for about 3 weeks, depositing a thin primary wall (Meinert and Delmer, (1984) Plant Physiol. 59: 1088-1097; Basra and Malik, (1984) Int Rev of Cytol 89: 65-113). At the time of transition to secondary wall cellulose synthesis, the fiber cells undergo a synchronous shift in the pattern of cortical microtubule and cell wall microfibril alignments, events which may be regulated upstream by the organization of actin (Seagull, (1990) Protoplasma 159: 44-59; and (1992) In: Proceedings of the Cotton Fiber Cellulose Conference, National Cotton Council of America, Memphis RN, pp 171-192.

Agrobacterium-mediated cotton transformation is described in Umbeck, United States Patents Nos. 5,004,863 and 5,159,135 and cotton transformation by particle bombardment is reported in WO 92/15675, published September 17, 1992. Transformation of *Brassica* has been described by Radke *et al.* (Theor. Appl. Genet. (1988) 75;685-694; Plant Cell Reports (1992) 11:499-505.

SUMMARY OF THE INVENTION

Novel DNA constructs and methods for their use are described which are capable of directing transcription of a gene of interest in cotton fiber, particularly early in fiber development and during secondary cell wall development. The novel constructs include a vector comprising a transcriptional and translational initiation region obtainable from a gene expressed in cotton fiber

and methods of using constructs including the vector for altering fiber phenotype. Both the endogenous 3' regions and 5' regions may be important in directing efficient transcription and translation.

Three promoters are provided from genes involved in the regulation of cotton fiber development. One, Rac13, is from a protein in cotton which codes for an animal Rac protein homolog. Rac13, shows highly-enhanced expression during fiber development. This pattern of expression correlates well with the timing of reorganization of the cytoskeleton, suggesting that the Rac13 cotton gene may, like its animal counterpart, be involved in the signal transduction pathway for cytoskeletal organization. Rac13 is a gene that is moderately expressed during fiber development turning on at 9 dpa and shutting down approximately 24 dpa. It is maximally expressed between 17-21 dpa developing fiber.

Another promoter from a cotton protein is designated 4-4. The 4-4 mRNA accumulates in fiber cells at day 17 post anthesis and continues towards fiber maturity, which occurs at 60 days or so post anthesis. Data demonstrates that the 4-4 promoter remains very active at day 35 post anthesis.

Also provided is a promoter from a lipid transfer protein (hereinafter sometimes referred to as "Ltp") which is preferentially expressed in cotton fiber.

The methods of the present invention include transfecting a host plant cell of interest with a transcription or expression cassette comprising a cotton fiber promoter and generating a plant which is grown to produce fiber having the desired phenotype.

Constructs and methods of the subject invention thus find use in modulation of endogenous fiber products, as well as production of exogenous products and in modifying the phenotype of fiber and fiber products. The constructs also find use as molecular probes. In particular, constructs and methods for use in gene expression in cotton embryo tissues are considered herein. By these methods, novel cotton plants and cotton plant parts, such as modified cotton fibers, may be obtained.

Also provided are constructs and methods of use relating to modification of color phenotype in cotton fiber. Such constructs contain sequences for expression of genes involved in the production of colored compounds, such as anthocyanins, melanin or indigo, and also may contain sequences which provide for targeting of the gene products to particular locations in the plant cell, such as plastid organelles, or vacuoles. Plastid targeting is of particular interest for expression of genes involved in aromatic amino acid biosynthesis pathways, while vacuolar targeting is of particular interest where the precursors required in synthesis of the pigment are present in vacuoles.

Of particular interest are plants producing fibers which are color, that is, with pigment produced in the fiber by the plant during fiber development, as opposed to fibers which are harvested and dyed or otherwise pigmented by separate processing. Fibers from a plant producing such colored fiber may be used to produce colored yarns and/or fabric which have not been subjected to any dyeing process. While naturally colored cotton has been available from various domesticated and wild type cotton varieties, the

instant application provides cotton fiber has a color produced by the expression of a genetically engineered protein.

Thus, the application provides constructs and methods of use relating to modification of color phenotype in cotton fiber. Such constructs contain sequences for expression of genes involved in the production of colored compounds, such as melanin or indigo, and also contain sequences which provide for targeting of the gene products to particular locations in the plant cell, such as plastid organelles, or vacuoles. Plastid targeting is of particular interest for expression of genes involved in the aromatic amino acid biosynthesis pathways, while vacuolar targeting is of particular interest where the precursors required in synthesis of the pigment are present in vacuoles.

DESCRIPTION OF THE DRAWINGS

Figure 1 shows the DNA sequence encoding the structural protein from cDNA 4-4.

Figure 2 shows the sequence to the promoter construct pCGN5606 made using genomic DNA from 4-4-6 genomic clone.

Figure 3 shows the sequence to the 4-4 promoter construct pCGN5610.

Figure 4 shows the cDNA sequence encoding the Rac13 gene expressed in cotton fiber.

Figure 5 shows the sequence the promoter region from the *rac*13 gene.

Figure 6 shows a restriction map for pCGN4735.

Figure 7 shows the sequence of the Ltp promoter region from a cotton fiber specific lipid transfer protein gene.

Figure 8 shows the arrangement of a binary vectors pCGN5148 and pCGN5616 for plant transformation to express genes for melanin synthesis and indigo synthesis, respectively.

Figure 9 provides the results of color measurements taken from fibers of the control Coker 130 cotton used in transformation using color constructs.

Figure 10 shows the results of measurements made of color of plants transformed by the pCGN5148 construct to express genes for melanin synthesis.

Figure 11 shows the results of measurements taken of the color of plants transformed by the pCGN5149 construct to express genes for melanin synthesis.

Figure 12 shows the results of measurements made of color of plants transformed to express genes for indigo synthesis, using construct pCGN5616.

Figure 13 shows control measurements made of naturally colored cotton plants which are produced by non-transgenic colored cotton plants.

DETAILED DESCRIPTION OF THE INVENTION

In accordance with the subject invention, novel constructs and methods are described, which may be used provide for transcription of a nucleotide sequence of interest in cells of a plant host, preferentially in cotton fiber cells to produce cotton fiber having an altered color phenotype.

Cotton fiber is a differentiated single epidermal cell of the outer integument of the ovule. It has four distinct growth phases; initiation, elongation (primary cell wall synthesis), secondary cell wall synthesis, and maturation. Initiation of fiber development appears to be triggered by hormones. The primary cell wall is laid down during the elongation phase, lasting up to 25 days postanthesis (DPA). Synthesis of the secondary wall commences prior to the cessation of the elongation phase and continues to approximately 40 DPA, forming a wall of almost pure cellulose.

The constructs for use in such cells may include several forms, depending upon the intended use of the construct. Thus, the constructs include vectors, transcriptional cassettes, expression cassettes and plasmids. The transcriptional and translational initiation region (also sometimes referred to as a "promoter,"), preferably comprises a transcriptional initiation regulatory region and a translational initiation regulatory region of untranslated 5' sequences, "ribosome binding sites," responsible for binding mRNA to ribosomes and translational initiation. It is preferred that all of the transcriptional and translational functional elements of the initiation control region are derived from or obtainable from the same gene. In some embodiments, the promoter will be modified by the addition of sequences, such as enhancers, or deletions of nonessential and/or undesired sequences. By "obtainable" is intended a promoter having a DNA sequence sufficiently similar to that of a native promoter to provide for the desired specificity of transcription

of a DNA sequence of interest. It includes natural and synthetic sequences as well as sequences which may be a combination of synthetic and natural sequences.

Cotton fiber transcriptional initiation regions chosen for cotton fiber modification may include the 4-4, *rac*13 and Ltp cotton fiber promoter regions provided herein.

A transcriptional cassette for transcription of a nucleotide sequence of interest in cotton fiber will include in the direction of transcription, the cotton fiber transcriptional initiation region, a DNA sequence of interest, and a transcriptional termination region functional in the plant cell. When the cassette provides for the transcription and translation of a DNA sequence of interest it is considered an expression cassette. One or more introns may be also be present.

Other sequences may also be present, including those encoding transit peptides and secretory leader sequences as desired.

Fiber-tissue transcription initiation regions of this invention are, preferably, not readily detectable in other plant tissues. Transcription initiation regions capable of initiating transcription in other plant tissues and/or at other stages of fiber development, in addition to the foregoing, are acceptable insofar as such regions provide a significant expression level in cotton fiber at the defined periods of interest and do not negatively interfere with the plant as a whole, and, in particular, do not interfere with the development of fiber and/or fiber-related parts.

Downstream from, and under the regulatory control of, the cotton fiber transcriptional/translational initiation control region is a nucleotide sequence of interest which provides for modification of the phenotype of fiber. The nucleotide sequence may be any open reading frame encoding a polypeptide of interest, for example, an enzyme, or a sequence complementary to a genomic sequence, where the genomic sequence may be an open reading frame, an intron, a noncoding leader sequence, or any other sequence where the complementary sequence inhibits transcription, messenger RNA processing, for example, splicing, or translation. The nucleotide sequences of this invention may be synthetic, naturally derived, or combinations thereof. Depending upon the nature of the DNA sequence of interest, it may be desirable to synthesize the sequence with plant preferred codons. The plant preferred codons may be determined from the codons of highest frequency in the proteins expressed in the largest amount in the particular plant species of interest. Phenotypic modification can be achieved by modulating production either of an endogenous transcription or translation product, for example as to the amount, relative distribution, or the like, or an exogenous transcription or translation product, for example to provide for a novel function or products in a transgenic host cell or tissue. Of particular interest are DNA sequences encoding expression products associated with the development of plant fiber, including genes involved in metabolism of cytokinins, auxins, ethylene, abscissic acid, and the like. Methods and compositions for modulating cytokinin expression are described in United States

Patent No. 5,177,307, which disclosure is hereby incorporated by reference. Alternatively, various genes, from sources including other eukaryotic or prokaryotic cells, including bacteria, such as those from *Agrobacterium tumefaciens* T-DNA auxin and cytokinin biosynthetic gene products, for example, and mammals, for example interferons, may be used.

Other phenotypic modifications include modification of the color of cotton fibers. Of interest are genes involved in production of melanin and genes involved in the production of indigo. Melanins are dark brown pigments found in animals, plants and microorganisms, any of which may serve as a source for sequences for insertion into the constructs of the present invention. Specific examples include the tyrosinase gene which can be cloned from *Streptomyces antibioticus*. The ORF438 encoded protein in *S. antibioticus* also is necessary for melanin production, and may provide a copper donor function. In addition, a tyrosinase gene can be isolated from any organism which makes melanin. The gene can be isolated from human hair, melanocytes or melanomas, cuttle fish and red roosters, among others. See, for example, EP Application No. 89118346.9 which discloses a process for producing melanins, their precursors and derivatives in microorganisms. Also, See, Bernan *et al.* Gene (1985) 37:101-110; and della-Cioppa *et al.* Bio/Technology (1990) 8:634-638.

Indigo may be obtained by use of genes encoding a mono-oxygenase such as xylene oxygenase which oxidizes toluene and xylene to (methyl) benzyl alcohol and also transforms indole to indigo. Cloning of the xylene oxygenase gene and the nucleotide

and amino acid sequences are described in unexamined Japanese Patent Application Kokai:2-119777, published May 7, 1990. A dioxygenase such as naphthalene dioxygenase which also converts indole to indigo finds use; the naphthalene dioxygenase gene nahA is described in Science (1983) 222: 167. For cloning, nucleotide sequence in characterization of genes encoding naphthalene dioxygenase of Pseudomonas putida. See, Kurkela *et al.* Gene (1988) 73:355-362. A tryptophanase gene sequence can be used in conjunction with an oxygenase to increase the amount of indole available for conversion to indigo. Sources of tryptophanase gene sequences include *E. coli* (see, for example, Deeley *et al.* (1982) *J. Bacteriol.* 151 :942-951).

Plastid targeting sequences (transit peptides) are available from a number of plant nuclear-encoded plastid proteins, such as the small subunit (SSU) of ribulose bisphosphate carboxylase, plant fatty acid biosynthesis related genes including acyl carrier protein (ACP), stearoyl-ACP desaturase, ß-ketoacyl-ACP synthase and acyl-ACP thioesterase, or LHCPII genes. The encoding sequence for a transit peptide which provides for transport to plastids may include all or a portion of the encoding sequence for a particular transit peptide, and may also contain portions of the mature protein encoding sequence associated with a particular transit peptide. There are numerous examples in the art of transit peptides which may be used to deliver a target protein into a plastid organelle. The particular transit peptide encoding sequence used in the instant invention is not critical, as long as delivery to the plastid is obtained.

As an alternative to using transit peptides to target pigment synthesis proteins to plastid organelles, the desired constructs may be used to transform the plastid genome directly. In this instance, promoters capable of providing for transcription of genes in plant plastids are desired. Of particular interest is the use of a T7 promoter to provide for high levels of transcription. Since plastids do not contain an appropriate polymerase for transcription from the T7 promoter, T7 polymerase may be expressed from a nuclear construct and targeted to plastids using transit peptides as described above. (See McBride *et al.* (1994) *Proc. Nat. Acad. Sci. 91*:7301-7305; see also copending US patent application entitled "Controlled Expression of Transgenic Constructs in Plant Plastids", serial no. 08/472,719, filed June 6, 1995, and copending US patent application SN 08/167,638, filed December 14, 1993 and PCT/US94/14574 filed December 12, 1994.) Tissue specific or developmentally regulated promoters may be useful for expression of the T7 polymerase in order to limit expression to the appropriate tissue or stage of development.

Targeting of melanin synthesis genes to vacuoles is also of interest in plant tissues which accumulate the tyrosine substrate involved in melanin synthesis in vacuoles. The protein signal for targeting to vacuoles may be provided from a plant gene which is normally transported across the rough endoplasmic reticulum, such as the 32 amino acid N-terminal region of the metallocarboxypeptidase inhibitor gene from tomato (Martineau *et al.* (1991) *Mol. Gen. Genet. 228* :281-286). In addition to the signal sequence, vacuolar targeting constructs also encode a

vacuolar localization signal (VLS) positioned at the carboxy terminus of the encoded protein. Appropriate signal sequences and VLS regions may be obtained from various other plant genes and may be similarly used in the constructs of this invention. Numerous vacuolar targetting peptides are known to the art, as are reviewed in Chrispeels *et al.*, Cell (1992) 68:613-616.

The Maize A1 gene which encodes a dihydroflavonol reductase, an enzyme of the anthocyanin pigmentation pathway is one such gene. In cells that express the A1 gene, dihydrokempferol is converted to 2-8 alkylleucopelargonidin, which may be further metabolized to pelargonidin pigment by endogenous plant enzymes. Other anthocyanin or flavonoid type pigments may also be of interest for modification of cotton cell fibers, and have been suggested for use in plant flowers (for a review of plant flower color, see van Tunen *et al.*, Plant Biotechnology Series, Volume 2 (1990) Developmental Regulation of Plant Gene Expression, D. Grierson ed.). Anthocyanin is produced by a progression of steps from cellular phenylalanine pools. The R anc C1 genes are maize regulatory proteins which are active by positively affecting upstream steps in the anthocyanin biosynthesis from these pools. The R gene is described in Perot and Cone (1989) Nucl. Acids Res., 17:8003, and the C1 gene is described in Paz-Ares *et al* (1987) EMBO, 6:3553-3558. Lloyd *et al.* (1992) Science, 258:1773-1775 discussed both genes.

Although cotton fibers in commercially grown varieties are primarily white in color, other naturally occurring cotton varieties have brown or reddish-brown fibers. Additionally, a

cotton line containing green colored fibers has been identified. Cotton lines providing such fibers are available from various sources, including the BC variety cottons (BC Cotton Inc., Box 8656, Bakersfield, CA 93389) and Fox Fibre cottons (Natural Cotton Colors, Inc., P.O. Box 791, Wasco, CA 93280).

The existence of such colored cotton lines suggests that the precursors required for the anthocyanin pigment pathways are present in cotton fibers cells, thus allowing further color phenotype modifications. Thus, the maize R and C1 genes could be used in enhancing the levels of of anthocyanin produced in fiber cells. As the R and C1 proteins are proteins with a positive control at the regulatory level on anthocyanin pigment precursor biosynthesis, these proteins are expressed in the nucleus, and not targetted to plastids or vacuoles.

For some applications, it is of interest to modify other aspects of the fiber. For example, it is of interest to modify various aspects of cotton fibers, such as strength or texture of a fiber. Thus, the appropriate gene may be inserted in the constructs of the invention, including genes for PHB biosynthesis (see, Peoples *et al. J. Biol. Chem.* (1989) 264: 15298-15303 and *Ibid.* 15293-15397; Saxena, *Plant Molecular Biology* (1990) 15:673-683, which discloses cloning and sequencing of the cellulose synthase catalytic subunit gene; and Bowen *et al. PNAS* (1992) 89:519-523 which discloses chitin synthase genes of *Saccharomyces cerevisiae* and *Candida albicans*. Various constructs and methods are disclosed for the use of hormones to effect changes to fiber quality in copending US patent application entitled "Cotton

Modification Using Ovary-Tissue Transcriptional factors", serial no. 08/397,652 filed February 2, 1995, the teachings of which are incorporated herein by reference.

Transcriptional cassettes may be used when the transcription of an anti-sense sequence is desired. When the expression of a polypeptide is desired, expression cassettes providing for transcription and translation of the DNA sequence of interest will be used. Various changes are of interest; these changes may include modulation (increase or decrease) of formation of particular saccharides, hormones, enzymes, or other biological parameters. These also include modifying the composition of the final fiber that is changing the ratio and/or amounts of water, solids, fiber or sugars. Other phenotypic properties of interest for modification include response to stress, organisms, herbicides, brushing, growth regulators, and the like. These results can be achieved by providing for reduction of expression of one or more endogenous products, particularly an enzyme or cofactor, either by producing a transcription product which is complementary (anti-sense) to the transcription product of a native gene, so as to inhibit the maturation and/or expression of the transcription product, or by providing for expression of a gene, either endogenous or exogenous, to be associated with the development of a plant fiber.

The termination region which is employed in the expression cassette will be primarily one of convenience, since the termination regions appear to be relatively interchangeable. The termination region may be native with the transcriptional

initiation region, may be native with the DNA sequence of interest, may be derived from another source. The termination region may be naturally occurring, or wholly or partially synthetic. Convenient termination regions are available from the Ti-plasmid of *A. tumefaciens,* such as the octopine synthase and nopaline synthase termination regions. In some embodiments, it may be desired to use the 3' termination region native to the cotton fiber transcription initiation region used in a particular construct.

As described herein, in some instances additional nucleotide sequences will be present in the constructs to provide for targeting of a particular gene product to specific cellular locations. For example, where coding sequences for synthesis of aromatic colored pigments are used in a construct, particularly coding sequences for enzymes which have as their substrates aromatic compounds such tyrosine and indole, it is preferable to include sequences which provide for delivery of the enzyme into plastids, such as an SSU transit peptide sequence. Also, for synthesis of pigments derived from tyrosine, such as melanin, targeting to the vacuole may provide for enhanced color modifications.

For melanin production, the tyrosinase and ORF438 genes from *Streptomyces antibioticus (Berman et al.* (1985) 37:101-110) are provided in cotton fiber cells for expression from a 4-4 and Rac13 promoter. In *Streptomyces,* the ORF438 and tyrosinase proteins are expressed from the same promoter region. For expression from constructs in a transgenic plant genome, the coding regions may be

provided under the regulatory control of separate promoter regions. The promoter regions may be the same or different for the two genes. Alternatively, coordinate expression of the two genes from a single plant promoter may be desired. Constructs for expression of the tyrosinase and ORF438 gene products from 4-4 and rac promoter regions are described in detail in the following examples. Additional promoters may also be desired, for example plant viral promoters, such as CaMV 35S, can be used for constitutive expression of one of the desired gene products, with the other gene product being expressed in cotton fiber tissues from the 4-4 and rac promoter.

Similarly, other constitutive promoters may also be useful in certain applications, for example the mas, Mac or DoubleMac, promoters described in United States Patent No. 5,106,739 and by Comai *et al.*, *Plant Mol. Biol.* (1990) *15:373-381)*. When plants comprising multiple gene constructs are desired, for example plants expressing the melanin genes, ORF438 and tyrosinase, the plants may be obtained by co-transformation with both constructs, or by transformation with individual constructs followed by plant breeding methods to obtain plants expressing both of the desired genes.

A variety of techniques are available and known to those skilled in the art for introduction of constructs into a plant cell host. These techniques include transfection with DNA employing *A. tumefaciens* or *A. rhizogenes* as the transfecting agent, protoplast fusion, injection, electroporation, particle acceleration, etc. For transformation with *Agrobacterium*,

plasmids can be prepared in *E. coli* which contain DNA homologous with the Ti-plasmid, particularly T-DNA. The plasmid may or may not be capable of replication in *Agrobacterium,* that is, it may or may not have a broad spectrum prokaryotic replication system such as does, for example, pRK290, depending in part upon whether the transcription cassette is to be integrated into the Ti-plasmid or to be retained on an independent plasmid. The *Agrobacterium* host will contain a plasmid having the *vir* genes necessary for transfer of the T-DNA to the plant cell and may or may not have the complete T-DNA. At least the right border and frequently both the right and left borders of the T-DNA of the Ti- or Ri-plasmids will be joined as flanking regions to the transcription construct. The use of T-DNA for transformation of plant cells has received extensive study and is amply described in EPA Serial No. 120,516, Hoekema, In: The Binary Plant Vector System Offset-drukkerij Kanters B.V., Alblasserdam, 1985, Chapter V, Knauf, *et al.*, Genetic Analysis of Host Range Expression by *Agrobacterium,* In: Molecular Genetics of the Bacteria-Plant Interaction, Puhler, A. ed., Springer-Verlag, NY, 1983, p. 245, and An, *et al.*, *EMBO J.* (1985) 4:277-284.

For infection, particle acceleration and electroporation, a disarmed Ti-plasmid lacking particularly the tumor genes found in the T-DNA region) may be introduced into the plant cell. By means of a helper plasmid, the construct may be transferred to the *A. tumefaciens* and the resulting transfected organism used for transfecting a plant cell; explants may be cultivated with transformed *A. tumefaciens* or *A. rhizogenes* to allow for transfer

of the transcription cassette to the plant cells. Alternatively, to enhance integration into the plant genome, terminal repeats of transposons may be used as borders in conjunction with a transposase. In this situation, expression of the transposase should be inducible, so that once the transcription construct is integrated into the genome, it should be relatively stably integrated. Transgenic plant cells are then placed in an appropriate selective medium for selection of transgenic cells which are then grown to callus, shoots grown and plantlets generated from the shoot by growing in rooting medium.

To confirm the presence of the transgenes in transgenic cells and plants, a Southern blot analysis can be performed using methods known to those skilled in the art. Expression products of the transgenes can be detected in any of a variety of ways, depending upon the nature of the product, and include immune assay, enzyme assay or visual inspection, for example to detect pigment formation in the appropriate plant part or cells. Once transgenic plants have been obtained, they may be grown to produce fiber having the desired phenotype. The fibers may be harvested, and/or the seed collected. The seed may serve as a source for growing additional plants having the desired characteristics. The terms transgenic plants and transgenic cells include plants and cells derived from either transgenic plants or transgenic cells.

The various sequences provided herein may be used as molecular probes for the isolation of other sequences which may be useful in the present invention, for example, to obtain related transcriptional initiation regions from the same or different

plant sources. Related transcriptional initiation regions obtainable from the sequences provided in this invention will show at least about 60% homology, and more preferred regions will demonstrate an even greater percentage of homology with the probes. Of particular importance is the ability to obtain related transcription initiation control regions having the timing and tissue parameters described herein. For example, using the probe 4-4 and rac, at least 7 additional clones, have been identified, but not further characterized. Thus, by employing the techniques described in this application, and other techniques known in the art (such as Maniatis, *et al.*, *Molecular Cloning,- A Laboratory Manual* (Cold Spring Harbor, New York) 1982), other transcription initiation regions capable of directing cotton fiber transcription as described in this invention may be determined. The constructs can also be used in conjunction with plant regeneration systems to obtain plant cells and plants; thus, the constructs may be used to modify the phenotype of fiber cells, to provide cotton fibers which are colored as the result of genetic engineering to heretofor unavailable hues and/or intensities.

Various varieties and lines of cotton may find use in the described methods. Cultivated cotton species include *Gossypium hirsutum* and *G. babadense* (extra-long stable, or Pima cotton), which evolved in the New World, and the Old World crops *G. herbaceum* and *G. arboreum*.

Color phenotypes can be assessed by the use of a colorimeter, an instrument which is already used to provide objective measurements of the color of cotton samples. A colorimeter uses a

combination of light sources and filters to make various estimates of a samples colors, sometimes referred to as tristimulus values.

In the past such estimtes have been used to calculate a value (Hunter's + b, described below) indicating the degree of yellowness of a cotton sample. The yellowness and reflectance (from Rd, the degree of lightness or darkness of the samples) has been used to provide cotton color measurements for grading. Tests are typically conducted by exposing the face of a sample to a controlled light source. A typical color chart showing how the official grade standards relate to Rd and+ b measurements is shown in Cotton, RJ Kohel and CF Lewis, Editors #24 in AGRONOMY Series- American Soc. Agromony (see Fig. 12-6).

Various colorimeter methods can be so used to quantify color and express it numerically. The Munsell method, devised by the American artist A.. Munsell, uses a classification system of paper color chips assorted according to their hue (Munsell Hue), lightness (Munsell Value), and saturation (Munsell Chroma) for visual comparison with a specimen color.

Other methods for expressing color numerically have been developed by an international organization concerned with light and color, the Commission Internationale de l'Eclairage (CIE), having a Central Bureau located at Kegelgasse 27, A-1030 Vienna, AUSTRIA. The two most widely known of these methods are the Yxy color space, devised in 1931 based on the tristimulus value XYZ, as defined by CIE, and the L*a*b* color space, devised in 1976 to provide more uniform color differences in relation to visual differences. Color spaces* such as these ar now used throughout

the world for color communication. The Hunter Lab color space was developed in 1948 by R.S. Hunter as a uniform color space which could be read directly from a photoelectric colorimeter (tristimulus method).

The L*C*h color space uses the same diagram as the L*a*b* color space, but uses cylindrical coordinates instead of rectangular coordinates. In this color space, L* indicates lightness and is the same as the L* of the L*a*b* color space, C* is chroma, and h is the hue angle. The value of chroma C is 0 at the center and increases according to the distance from the center. Hue angle is defined as starting at the +a axis of the L*a*b* space, and is expressed in degrees in a counterclockwise rotation. Thus, relative to the L*a*b* space, 0° and 360° would be at the +a* line, 90° would be +b*, 180° would be -a* and 270° would be -b*.

All of the above methods can be used to obtain precise measurements of a cotton fiber color phenotype.

EXPERIMENTAL

The following examples are offered by way of illustration and not by limitation.

Example 1

cDNA libraries

Tissue preparation for cDNA synthesis

Leaf and root tissue were isolated from 8 inch tall greenhouse grown seedlings and immediately frozen in liquid nitrogen. Flowers were collected at the rapidly expanding 3 day

preanthesis stage and also frozen. Seed was collected from 21 day postanthesis locules which had been removed from the boll and frozen entire in liquid nitrogen. Once frozen, the fiber was removed from the seed and the denuded seed used for RNA isolation. All fibers were removed from the seed under liquid nitrogen and the fiber was ground to a powder prior to RNA isolation. Fibers were from bolls which had been tagged at anthesis.

DNA and RNA Manipulations

The lambda ZapII™ cDNA library system of Stratagene was used for screening, and was prepared from cDNA derived from poly-A$^+$ mRNA isolated from fibers of *Gossypium hirsutum* cultivar Acala SJ-2. The fibers were isolated from bolls harvested at approximately 21 dpa using field-grown plants in Israel.

Total RNA was isolated from 21 dpa seeds (*G. hirsutum* cv Coker 130 from which the fiber had been removed) using the method of Hughes and Galau ((1988) Plant Mol Biol Reporter, 6:253-257.) All other RNAs were prepared according to Hall *et al.* ((1978), Proc Natl Acad Sci USA 75: 3196-3200), with the following modifications. After the second 2M LiCl wash, the pellet was dissolved in 1/10 original volume of 10 mM Tris pH7.5 and brought to 35mM potassium acetate pH6.5 and 1/2 volume EtOH was added slowly. The mixture was placed on ice for 15 minutes and then centrifuged at 20,000 x *g* for 15 minutes at 4^{O}C. The potassium acetate concentration was brought to 0.2M, 2 1/2 volumes EtOH added and the RNA placed at -20^{O}C for several hours. The precipitate was centrifuged at 12,000 x *g* for 30 minutes at 4^{O}C

and the pellet was resuspended in diethylpyrocarbonate-treated water. Poly-A^+ RNA was prepared from total mRNA utilizing an oligo(dT)-cellulose kit (Becton Dickenson) and following the manufacturer's protocol.

Cotton genomic DNA was prepared as follows. Four grams of young cotton leaf tissue (cv Coker 130) was ground to a powder in N_2 and placed in an Oak Ridge tube with 0.4g polyvinylpyrolidone and 20mls extraction buffer (200mM Ches/NaOH ph9.1, 200mM NaCl, 100mMEDTA/NaOH pH9.0, 2% SDS, 0.5% Na deoxycholate, 2% Nonidet NP-40, 20mM B-mercaptoethanol) was added to sample, gently mixed and incubated at 65^{O}C in a shaking water bath for 10 minutes. 7.0 mls of 5M potassium acetate pH6.5 was added and carefully mixed. Incubation was carried out on ice for 30 minutes with gentle mixing every 5 minutes. The sample was centrifuged for 20 minutes at 21,000 x g and the supernatant was filtered through Miracloth into another tube and centrifuged as before. The supernatant was again filtered through Miracloth into 15 mls of room temperature isopropanol in an Oak Ridge tube. After gentle mixing, the sample was incubated at room temperature for 10-60 minutes until the DNA precipitated. The DNA was spooled and allowed to air dry before being resuspended in 4 mls of TE on ice for 1 hour. CsCl was added to 0.97g/ml final concentration and 300 ul 10mg/ml ethidium bromide was also added before filling VTi80 quick seal tubes. The sample was centrifuged overnight at 225,000 x g overnight. The DNA was extracted with water saturated butanol and enough water was added to bring the volume to 4 mls before adding 2 volumes

EtOH. The DNA was spooled, air dried and resuspended in 200 ul sterile water.

Northern and Southern Analysis

For Northerns, 10ug of total RNA was isolated from various tissues, separated by electrophoresis in 1.2% agarose-formaldehyde gels and transfered onto Nytran Plus membranes (Schleicher and Schuell). Hybridization conditions consisted of a solution containing 50% formamide(v/v), 5xSSC, 0.1% SDS, 5mM EDTA, 10X Denhardts solution, 25mM sodium phosphate pH6.5 and 250 ug/ml carrier DNA. Washes were performed in 2xSSC, 0.1% SDS at $42^{O}C$ 3 times for 30 minutes each time.

Cotton genomic DNA (12ug) was digested with various restriction endonucleases, electrophoresed in 0.9% agarose gels and blotted onto Nytran Plus membranes. Hybridization and filter washing conditions for both the 3' specific and full-length cDNA insert probes were as described for Northern analysis.

Probes derived from 3'-untranslated regions were synthesized via oligonucleotide primers from the Rac13 cDNA, corresponding to bases 600-619 and 843-864 (Figure 4). Each set of primers was used in a polymerase chain reaction to synthesize copies of 3'-specific DNA sequences. These sequences were used as templates in the generation of single-stranded, ^{32}P-labeled probes off the antisense strand in a polymerase chain reaction. The full-length cDNA inserts for Rac13 were used as templates for double stranded, random primed probes using the Prime-It kit (Stratagene).

Example 2

Isolation of cDNA Clones from Cotton

cDNA to the 4-4 clone was isolated from the cotton fiber library described above, and shown to express in fiber but not other tissues. This sequence was not related to any known protein. Only 400 kb of encoding sequence was present in this clone, so the library was rescreened using the cDNA to obtain full-length clones. The full-length encoding sequence is provided in Figure 1.

By comparing sequences of random cDNA clones against various sequence data banks via BLAST, a National Center for Biotechnology Information service, a clone, designated #105, was found to have an encoding sequence related to that of a reported lipid transfer protein.

Another clone was sequenced which showed high homology to animal Rac proteins. This clone, designated Rac, was not quite full-length, and the library was re-screened using this initial Rac DNA segment as probe. Of approximately 130,000 primary plaques screened, 56 screened positive; of these, 14 clones were isolated and sequenced. Of these 14 clones, 12 showed identical sequence homology to the original Rac clone and one of these cDNA clones encoded a full length cDNA and received the name Rac13. Figure 4 shows the cDNA sequence encoding the Rac13 gene expressed in cotton fiber.

One other partial-length cDNA clone, designated Rac9, was clearly related, but distinct in DNA and amino acid sequence from Rac13. Re-screening of 150,000 plaques resulted in the isolation

of 36 positive clones of which only two clones corresponded to the Rac9 sequence (both full-length clones), the remainder being Rac13. These results suggest that cotton contains genes for at least two distinct Rac proteins. Based upon the frequency of clone isolation, Rac13 is relatively highly-expressed and Rac9 less so in cotton fibers at 21 days post-anthesis (dpa), the age at which polyA$^+$ mRNA was isolated for library construction.

Comparisons of the deduced amino acid sequence of Rac13 with other small G-proteins showed that the cotton Rac proteins are very closely related to the Rho1 protein sequence deduced from a cDNA clone isolated recently from pea (Yang and Watson, *supra*). After the pea Rho1, mammalian Rac proteins show the highest homology with the cotton Rac proteins. Other proteins of the *rho* subfamily, such as the yeast CDC42 and human RhoA, are also clearly related to the cotton Rac genes. By contrast, the other small G-proteins of the *Rab/YPT* subfamily isolated from plants such as the example shown of the tobacco RAB5 protein, as well as the human Ras proteins, are least homologous to the cotton Rac proteins of all the small G-proteins compared. The cotton and pea proteins, as well as the mammalian Racs, all have pI's above 9, whereas those of other *rho* and *ras* proteins are in the range of 5.0-6.5.

Example 3

<u>Expression of Cotton Fiber Genes in Developing Fibers</u>

Expression of the Rac13 and 4-4 genes was assessed using mRNA prepared from various cotton tissues and from fibers at

different stages of development. Blots were hybridized with probes derived from untranslated regions of Ltp, Rac13 and 4-4 genes. The gene for Rac13 exhibits highly-enhanced expression in fibers; virtually no detectable mRNA is present in leaves, roots, or flower parts, even under conditions of extended development time. Rac13 expression is detected in seeds at an age that corresponds to the highest expression levels observed in fiber tissue derived from seeds of this same age. The pattern of Rac13 expression in fibers is very dependent upon the developmental stage. Expression is very low during the stage of primary wall synthesis (0-14 dpa, see Meinert and Delmer, 1977), reaches a maximum during the transition to secondary wall synthesis (about 15-18 dpa), and declining during the stage of maximal secondary wall cellulose synthesis (about 24-28 dpa).

4-4 mRNA is begins to accumulate in fiber cells only at day 17 post anthesis and continues through at least day 35 post anthesis. Levels peak at day 21 and remain high. 4-4 mRNA is not detected in other cotton tissues, and is not detected in fiber tissue before onset at 17 days post anthesis.

The #105 lipid transfer protein cDNA clone was used as a probe against cotton tissue and in a cotton fiber northern. The northern showed that the cotton fiber Ltp is highly expressed in cotton fiber. The mRNA that codes for this protein is expressed throughout fiber development at extremely high level. Northern blots indicate that this mRNA is expressed at 5 dpa and is continually expressed at a high level at 40 dpa.

Example 4

Genomic DNA

cDNA for both the 4-4 and Rac13 was used to probe for genomic clones. For both, full length genomic DNA was obtained from a library made using the lambda dash 2 vector from Stratagene™, which was used to construct a genomic DNA library from cotton variety Coker 130 (*Gossypium hirsutum* cv. coker 130), using DNA obtained from germinating seedlings.

The cotton genomic library was probed with a 3'-specific Ltp probe and 6 genomic phage candidates were identified and purified. Figure 7 provides an approximately 2 kb sequence of the Ltp promoter region which is immediately 5' to the Ltp encoding region.

Six genomic phage clones from the cotton genomic library were identified using a 3'-specific probe for the Ltp mRNA. This was done to select the promoter from the Ltp gene that is maximally expressed in cotton fiber from the family of Ltp genes in cotton. The Ltp promoter is active throughout the fiber development period.

Example 5

Preparation of 4-4 Promoter Constructs

pCGN5606

The pCGN5606 promoter construct comprises the 4-4 cotton fiber expression cassette in a first version, version I (Figure 2). The sequences from nt1 to 65 and nt 5,494 to 5,547 correspond to fragments of the pBluescriptII polylinker where this cassette

is cloned. Unique restriction enzyme sites present in these regions flanking the cassette allow the cloning of the fiber expression cassette into binary vectors including the pCGN 5138 and 1547 series.

The sequences from nt57 to 5,494 are contained in a lambda phage clone of a cotton Coker 130 genomic library. This lambda genomic clone was given the designation 4-4(6).

The region from nt 65 to nt 4,163 corresponds to the 5' flanking region of the 4-4(6) gene. At nt 4,163 there is a NcoI restriction site sequence that corresponds to the first codon of the 4-4 (6)ORF.

The region from nucleotide 4,163 to 4,502 corresponds to part of the 4-4 (6)ORF. The sequence from nt 4,502 to 4,555 is a synthetic polylinker oligonucleotide that contains unique target sites for the restriction enzymes EcoRI, SmaI, SalI, NheI and BglII. This fragment from nt4,163 to 4,555 is a stuffer fragment and is left in place to facilitate the monitoring of cloning manipulations.

The genes to be expressed in cotton fiber cells using this cassette can be cloned between the NcoI restriction site and any of the polylinker sites. This operation will replace the stuffer fragment with the gene of interest. The region from nt 4,555to 5,494 corresponds to the 940 nucleotides downstream of the stop codon and constitute the 3' flanking region of the 4-4 (6) gene. There is a unique AscI restriction enzyme site at nt 5483.

<u>pCGN5610</u>

The pCGN5610 construct is a second version of a 4-4 cotton fiber expression cassette, version II, which is a modified version of pCGN5606. The two versions of the 4-4 cotton fiber expression cassette are designed to allow the cloning of tandem arrays of two fiber cassettes in one binary plasmid. The differences with respect to pCGN5606 are very minor and described below.

The XbaI restriction site in the region of nt 1 to 65 has been deleted by standard cloning manipulations.

The polylinker region is in the reverse orientation of pCGN5606. There is a unique XbaI restriction enzyme site at nt5484. The sequences from nt1 to 57 and nt 5,494 to 5,518 of pCGN5610 correspond to fragments of the pBluescriptII polylinker where this cassette is cloned. Unique restriction enzyme sites present in these regions allow the cloning of the fiber expression cassette into binary vectors of the pCGN 5138 and 1547 series.

The sequences from nt57 to 5,494 are contained a lambda phage clone of a Coker 130 genomic library. This clone is described in my notebook as lambda genomic clone 4-4(6). The region from nt 57 to nt 4,155 corresponds to the 5' flanking region. At nt 4,155 there is a NcoI restriction site sequence that corresponds to the first codon of the 4-4 ORF.

The region from nucleotide 4,156 to 4,500 corresponds to part of the 4-4 ORF. This fragment from nt4,156 to 4,550 is a stuffer fragment and is left in place to facilitate the monitoring of cloning manipulations. The sequence from nt 4,500 to 4,550 is a synthetic polylinker oligonucleotide containing unique target

sites for the restriction enzymes BglII, NheI, SalI, SmaI and EcoRI.

The genes to be expressed in cotton fiber cells using this cassette can be cloned between the NcoI restriction site and any of the polylinker sites. This operation replaces the stuffer fragment with the gene of interest. The region from nt 4,550 to 5,494 corresponds to the 940 nucleotides downstream of the stop codon and constitute the 3' flanking region of the 4-4 (6) gene.

Example 6

Preparation of Rac13 Promoter Constructs

Genomic clone

From a genomic clone designated 15-1, mapping was done with restriction endonucleases. The largest fragment with the Rac13 coding region was identified. Theis was a Pst fragment, and when subcloned in the Bluescript™ KS+ vector (BSKS+; Stratagene) was named pCGN4722. The insert had a length of 9.2 kb.

The region of the Pst fragment with the Rac13 coding sequence was identified. DNA sequence was determined for approximately 1.7 kb 5' of the start codon and approximately 1.2 kb 3' of the stop codon. The entire Rac coding region (exons and introns) was conveniently flanked by Nde1 sites.

pCGN4722 was digested with Xba1, and a 2.7 kb fragment was removed. Religation gave pCGN4730, which was then digested with Nde1, dropping out a 1.7 kb fragment containing the entire Rac coding region. Religation yielded pCGN4731.

A polylinker region was created using overlapping synthetic oligonucleotides which were PCR'ed using primers homologous to the 5' and 3' ends of the resynthesized section. The resulting product was digested with EcooR1 and Hind III and ligated into BSKS+ at the EcoR1 and Hind III sites. The resulting plasmid was designated pCGN4733.

pCGN4731 and pCGN4633 were digested with Nde1 and the Nde1 fragment containing the synthesized polylinker region from pCGN4733 was dropped in the Nde1 site of 4731, giving pCGN4734. This last plasmid was digested with Sal and Xba, and so was pCGN5133. pCGN5133 was the 9.2 kb pst fragment in BSKS+ where the polylinker sites flanking the insert were altered to different sites for ease of manipulation. The fragment from pCGN4734 was then placed into the equivalent site of pCGN5143, giving pCGN4735.

A sequence for approximately 3 kb of the promoter construct pCGN4735 is provided in Figure 5. The resynthesized sequence falls between the Nde1 sites located at bases 1706 and 1898 of the sequences. Thus, the sequence in Figure 5 includes approximately 1.7 kb 5' to the Nde1 site 5' to the resynthesized polylinker region. There is a roughly 2.5 kb sequence 5' from this sequence which is not provided in Figure 5, relative to the total 9.2 kb insert. The sequence of Figure 5 also includes approximately 1.1 kb 3' to the 3' Nde1 site. Approximately 3 kb which is most 3' in the Rac13 insert is not provided in Figure 5. A map for pCGN4735 is provided in Figure 6.

Example 7

Pigment Synthesis Genes

Melanin

A binary construct for plant transformation to express genes for melanin synthesis is prepared as follows. The melanin genes were originally isolated from the common soil bacterium *Streptomyces antibioticus* (Bernan et al. (1985) 34:101-110). Melanin production is composed of a two gene system. The first gene, *tyrA*, encodes the catalytic unit responsible for the polymerization of the amino acid tyrosine, the primary substrate, and is termed tyrosinase. The second gene, ORF438, is responsible for binding copper and delivering copper to the tyrosinase and activating the enzyme. Expression of both the ORF438 and *tyrA* genes ensures maximal tyrosinase activity.

The genes for both ORF438 and *tyrA* were fully re-synthesized with respect to their DNA sequence. This was performed as the initial DNA sequence isolated from *Streptomyces* has a very high guanine and cytosine (G+C) DNA content. Thus, the ORF438 and *tryA* genes were re-synthesized to appear more "plant-like" (reduced G+C content) with respect to plant preferred codons encoding their corresponding amino acids.

Indigo

Indigo production involves conversion of the amino acid tryptophan, the primary substrate, into indole which is then converted into indoxyl. Molecules of indoxyl spontaneously convert to indigo in the presence of oxygen. A two gene system was used to affect indigo production in fiber cells. The first

gene (*tna*) was obtained from the bacterium *E. coli* and encodes the enzyme tryptophanase. The designation *tna* stands for the gene encoding tryptophanase from *E. coli*, an enzyme which converts tryptophan to indole (Stewart et al., (1986) *J Bacteriol* 166:217-223).

The *pig* designation is used for the encoding sequence to the protein for indigo production from *Rhodococcus*, which produces indigo from indole (Hart *et al.*, (1990) J Gen Microbiol 136:1357-1363). Both *tna* and *pig* were obtained by PCR. Tryptophanase is responsible for the conversion of tryptophan to indole, while the second gene (pig) encodes an indole oxygenase enzyme responsible for the conversion of indole to indoxyl. Both these bacterial genes were utilized in their native form.

Example 8

Constructs for Targeting Pigment Synthesis Genes

For plastid targeting, the constructs contain a fragment of the tobacco ribulose bisphosphate carboxylase small subunit gene encoding the transit peptide and 12 amino acids of the mature protein (Tssu) positioned in reading frame with the appropriate encoding sequence.

For vacuolar targeting of the melanin synthesis genes, constructs include a fragment of the metallocarboxypeptidase inhibitor gene, encoding the entire 32 amino acid N-terminus signal peptide of that protein plus 6 amino acids of the mature protein (CPI+6) (Martineau *et al., supra)*, positioned in reading frame with the appropriate encoding sequences. In addition to the

signal peptide, a sequence encoding a vacuolar localization signal (VLS) is inserted 3' of the protein encoding sequence.

Constructs which contain encoding sequences for bacterial genes involved in biosynthesis of pigmented compounds and sequences for directing transport of the encoded proteins into plastids or vacuoles are prepared as follows.

Melanin

The re-synthesized ORF438 and *tyrA* genes were treated in two distinct ways depending on which compartment in the fiber cell the final protein products would be localized. One chimeric gene/plant binary construct (designated pCGN5148) contained the genes targeted to the fiber cell plastids. To do this, 12 amino acids of a gene for the small subunit of carboxylase (SSU) plus the original 54 amino acid SSU transit peptide were fused to the amino termini of both the ORF438 and *tyrA* gene products respectively. These peptide sequences allow the ORF438 and *tyrA* gene products (proteins) to be efficiently targeted to the plastid. This targeting was initiated as the plastid is the site of tyrosine production within the fiber cell.

The second chimeric gene/plant binary construct (designated pCGN5149) contained the ORF438 and *tyrA* genes targeted to the vacuole within the fiber cell. Based on information from other biological systems, it was postulated that the fiber cell vacuole may contain a high concentration of tyrosine for melanin polymerization. Both the ORF438 and *tryA* genes contain the 29 amino acid signal peptide from a tomato carboxypeptidase inhibitor

(CPI) protein as amino terminal gene fusions to direct these proteins to the endoplasmic reticulum (ER) secretory system of the fiber cell.

In addition, the *tyrA* gene has an 8 amino acid vacuolar targeting peptide (VTP) from CPI fused at the carboxy terminus so that the mature copper-activated tyrosinase will eventually be targeted to the vacuole of the fiber cell. Both the ORF438 and *tyrA* proteins also had potential glycosylation sites removed via site-directed mutagenesis of the ORF438 and *tyrA* genes respectively. Potential plant cell glycosylation of these proteins upon their expression in fiber cells could result in tyrosinase inactivation, hence removal of potential glycosylation sites was deemed necessary.

Indigo

The only modification to the indigo genes was the fusion of the tobacco SSU transit peptide encoding DNA sequences onto the amino terminal region of both the *tna* and pig genes to affect the localization of both the tryptophanase and indole oxygenase proteins to the fiber cell plastid. These are the same exact gene fusions that were made for the plastid-directed proteins for melanin production in construct 5148. The *tna* and pig gene products were targeted to the fiber cell plastid as that is the primary site of tryptophan synthesis.

Example 9

Expression Constructs

Melanin

The modified genes for both the plastid and vacuolar targeted ORF438 and tyrosinase proteins were placed into a fiber expression cassette to be "switched" on during development of the cotton fiber cell. The "switch" (promoter) utilized for the melanin constructs was 4-4. The modified ORF438 and *tyrA* genes were cloned into the 4-4 promoter cassette and these chimeric genes then inserted into a binary plasmid to create plasmids pCGN5148 and pCGN5149, containing the modified genes for plastid and vacuolar targeted ORF438 and tyrosinase proteins, respectively. These binary plasmids also contain genetic determinants for their stable maintenance in *E. coli* and *Agrobacterium* and also contain a chimeric gene for plant cell expression of the bacterial kanamycin resistance gene. This kanamycin resistance marker allows for the selection of transformed versus non-transformed cotton cells when plant hypocotyl or leaf segments are infected with Agrobacterium containing the binary plasmids.

A block diagram of the plasmid pCGN5149, having vacuolor targetting sequences, is shown in Figure 8. Plasmid pCGN5148 (not shown) is constructed the same as 5149, only pCGN5148 has plastid-targetting sequences.

Indigo

As with the melanin genes, the plastid-directed *tna* and pig genes were placed in the fiber-specific 4-4 promoter cassette and these chimeric genes subsequently inserted into a binary plasmid

to create plasmid pCGN5616. A block diagram of plasmid pCGN5616 is shown in Figure 8.

Anthocyanin

A construct has been prepared for the expression of the maize R and CI genes in developing cotton fiber. These genes are known to be responsible for the production of Anthocyanin pigments by acting in a regulatory manner to turn on the chalcone pathway for production of anthocyanins (red spectrum colors). The R and CI genes were placed under the control of the Rac13 promoter cassette. A binary plasmid designated pCGN4745 (not shown), contains both the R and CI genes each under control of the Rac13 promoter.

Example 10

Cotton Transformation

Explant Preparation

Coker 315 seeds are surface disinfected by placing in 50% Clorox (2.5% sodium hypochlorite solution) for 20 minutes and rinsing 3 times in sterile distilled water. Following surface sterilization, seeds are germinated in 25 x 150 sterile tubes containing 25 mls 1/2 x MS salts: 1/2 x B5 vitamins: 1.5% glucose: 0.3% gelrite. Seedlings are germinated in the dark at 28°C for 7 days. On the seventh day seedlings are placed in the light at 28±2°C.

Cocultivation and Plant Regeneration

Single colonies of *A. tumefaciens* strain 2760 containing binary plasmids pCGN2917 and pCGN2926 are transferred to 5 ml of MG/L broth and grown overnight at 30°C. Bacteria cultures are diluted to 1 x 10^8 cells/ml with MG/L just prior to cocultivation. Hypocotyls are excised from eight day old seedlings, cut into 0.5-0.7 cm sections and placed onto tobacco feeder plates (Horsch *et al.* 1985). Feeder plates are prepared one day before use by plating 1.0 ml tobacco suspension culture onto a petri plate containing Callus Initiation Medium CIM without antibiotics (MS salts: B5 vitamins: 3 % glucose: 0.1 mg/L 2,4-D: 0.1 mg/L kinetin: 0.3% gelrite, pH adjusted to 5.8 prior to autoclaving). A sterile filter paper disc (Whatman #1) was placed on top of the feeder cells prior to use. After all sections are prepared, each section was dipped into an *A. tumefaciens* culture, blotted on sterile paper towels and returned to the tobacco feeder plates.

Following two days of cocultivation on the feeder plates, hypocotyl sections are placed on fresh Callus Initiation Medium containing 75 mg/L kanamycin and 500 mg/L carbenicillin. Tissue was incubated at 28±2°C, 30uE 16:8 light:dark period for 4 weeks. At four weeks the entire explant was transferred to fresh callus initiation medium containing antibiotics. After two weeks on the second pass, the callus was removed from the explants and split between Callus Initiation Medium and Regeneration Medium (MS salts: 40mM KNO_3: 10 mM NH_4Cl:B5 vitamins:3% glucose:0.3% gelrite:400 mg/L carb:75 mg/L kanamycin).

Embryogenic callus was identified 2-6 months following initiation and was subcultured onto fresh regeneration medium.

Embryos are selected for germination, placed in static liquid Embryo Pulsing Medium (Stewart and Hsu medium: 0.01 mg/l NAA: 0.01 mg/L kinetin: 0.2 mg/L GA3) and incubated overnight at 30°C. The embryos are blotted on paper towels and placed into Magenta boxes containing 40 mls of Stewart and Hsu medium solidified with Gelrite. Germinating embryos are maintained at 28±2°C 50 uE $m^{-2}s^{-1}$ 16:8 photoperiod. Rooted plantlets are transferred to soil and established in the greenhouse.

Cotton growth conditions in growth chambers are as follows: 16 hour photoperiod, temperature of approximately 80-85°, light intensity of approximately 500µEinsteins. Cotton growth conditions in greenhouses are as follows: 14-16 hour photoperiod with light intensity of at least 400µEinsteins, day temperature 90-95°F, night temperature 70-75°F, relative humidity to approximately 80%.

Plant Analysis

Flowers from greenhouse grown T1 plants are tagged at anthesis in the greenhouse. Squares (cotton flower buds), flowers, bolls etc. are harvested from these plants at various stages of development and assayed for enzyme activity. GUS fluorometric and histochemical assays are performed on hand cut sections as described in co-pending application filed for Martineau *et al.*, *supra*. For fiber color characteristics, plants are visually inspected, or northern or western analysis can be performed, if necessary.

Example 11

Expression of Transgenic Pigment Synthesis Genes

Melanin

Plants that exhibited resistance to the kanamycin selectable marker via a leaf assay and corresponding Western analysis were considered transformed. Transgenic fiber was collected from individual plant transformants at different stages of fiber development and analyze in two ways. One was to analyze fiber at a single developmental time point for each transgenic cotton plant to compare tyrosinase expression between transgenic events. The other was to screen developing fiber from selected plants to analyze the timing of tyrosinase expression under the control of the fiber-specific 4-4 promoter, by Western blots using antisera prepared against purified tyrosinase protein.

For the plastid-targeted construct pCGN5148 9 of 13 events screened for tyrosinase expression were positive, while 13 of the 16 transformed vacuolar-targeted construct pCGN5149 events which were screened were positive. Expression level in the fiber in tyrosinase positive plants is approximately 0.1-0.5% fiber cell protein. Clearly, the cotton fiber cells comprising the DNA color constructs DNA produce the necessary proteins required for synthesis of a pigment.

Visually, the lint from the tyrosinase positive events exhibits color to varying degrees, while plants that do not express the enzyme do not exhibit any color. Colorimeter measurements of cotton fiber taken from control Coker 130 plants

and plants from various events transformed with pCGN5148 are provided in Figures 9 and 10, respectively.

Fiber from pCGN5148 (plastid-directed) plants demonstrates a bluish-green color phenotype. One event, 5148-50-2-1 included cotton fiber cells (linters) which were colored and which had an negative a* value less than - 8.0, as measured on the L*a*b* color space. Coker 130 cotton fiber cells do not typically demonstrate a negative a* value.

These colored cotton cells also had a color located on the L*C*h color space with a relatively high hue angle value h, greater than 135°. Normal Coker 130 fibers have a similar value which is not greater than about 90° as measured by this method.

Results of colorimeter measurements of cotton fiber taken from plants transformed with pCGN5149 are provided in Figure 11. Fiber from plants expressing tyrosinase from construct pCGN5149 (vacuolar-targetted) tends to have a light brown phenotype.

Indigo

Resistance to the kanamycin selectable marker via leaf assay and Western analysis was again the criterion for designating a plant as transformed by pCGN5616. Transgenic fiber was collected from individual plant transformants at different stages of fiber development. The transgenic developing fiber is screened from selected plants to analyze the timing of *tna* and pig gene expression under the control of the fiber-specific 4-4 promoter and fiber is also analyzed at a single developmental time point for each transgenic cotton plant for comparison of both

tryptophanase and indole oxygenase expression between transgenic events, by using Western blots with antisera prepared against the tryptophanase and indole oxygenase proteins.

For the indigo events, 15 of 24 screened plants were positive for expression of both the tryptophanase and indole oxygenase enzymes. Expression levels in the fiber of these proteins is between 0.05-0.5% fiber cell protein. Approximately half of these transformants are expressing both genes in the fiber resulting in a very faint light blue color phenotype. Visually, there is a faint blue color in the majority of these positive events, particularly in 20-30 dpa fiber in the unopened boll. Results of colorimeter measurements of cotton fiber taken from various events of plants transformed with pCGN5616 are provided in Figure 12. Many of these events had relatively low a* values (less than 2) with elevated b* values (greater than 10), as measured on the L*a*b* color space. Similarly, several 5149 events also measured with an a* value less than 2 while maintaining a b* value greater than 10.

BC Cotton

Colorimeter measurements taken on naturally colored fiber from four separate BC cotton lines is provided in Figure 13.

The above results demonstrate that the color phenotype of a transgenic cotton fiber cell can be altered by expressing pigment synthesis genes. The transgenic cotton fiber cells include both a pigment synthesizing protein, and pigment produced by the pigment

synthesizing protein. As shown from the results of Figures 9 through 13, expression of a pigment gene of interest can result in cotton fiber cells in which the synthesis of pigments combined with appropriate targeting sequences results in modification of color phenotype in the selected plant tissue, yielding colored cotton fiber by expression from a genetically engineered construct.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application are specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail, by way of illustration and example for purposes of clarity and understanding, it will be readily apparent to those of ordinary skill in the art that certain changes and modifications may be made thereto, without departing from the spirit or scope of the appended claims.

CLAIMS

What is claimed is:

1. A DNA construct comprising as operably joined components in the direction of transcription, a cotton fiber transcriptional factor and an open reading frame encoding a protein of interest, wherein said transcriptional factor is selected from the group consisting of the Ltp, the 4-4 and the *rac* promoter sequences.

2. The DNA construct according to Claim 1, further comprising a transport signal encoding sequence from a plant nuclear-encoded gene.

3. The DNA construct according to Claim 2, wherein said transport signal encoding sequence comprises a plastid transit peptid.

4. The DNA construct according to Claim 1, wherein said transport signal encoding sequence encodes a signal peptide which provides for transport across the rough endoplasmic reticulum.

5. The DNA construct according to Claim 4, wherein said sequence further comprises, 3' to said open reading frame, a vacuolar localization signal.

6. The DNA construct of Claim 1 wherein said pigment is melanin or indigo.

7. The DNA construct of Claim 6 wherein said open reading frame is from a bacterial gene.

8. The DNA construct of Claim 7 wherein said bacterial gene is selected from the group consisting of ORF438, *tyrA*, anthocyanin R gene, anthocyanin C1 gene, *pig*, and *tna*.

9. A plant cell comprising a DNA construct of Claim 1.

10. A cotton plant cell according to Claim 9.

11. A cotton fiber cell according to Claim 10.

12. A plant comprising a cell of any one of Claims 9-11.

13. A method of modifying fiber phenotype in a cotton plant, said method comprising:

transforming a plant cell with DNA comprising a construct for expression of a protein in a pigment biosynthesis pathway, wherein said construct comprises as operably joined components:

a transcriptional initiation region functional in cells of said cotton plant,

an open reading frame encoding a protein of interest, and

a transcriptional termination region functional in cells of said cotton plant,

wherein said plant cell comprises a substrate of said protein; and

growing said plant cell to produce a cotton plant, wherein said protein reacts with said substrate to produce said pigment.

14. The method of Claim 13 wherein said construct further comprises a transport signal encoding sequence from a plant nuclear-encoded gene.

15. The method of Claim 13 wherein said transport signal encoding sequence encodes a signal peptide which provides for transport across the rough endoplasmic reticulum.

16. The method of Claim 13 wherein said DNA comprises constructs for expression of two proteins in a pigment biosynthesis pathway, wherein each of said constructs comprises components i) through iv), and wherein said two proteins are not encoded by the same gene.

17. The method of Claim 16 wherein said pigment is melanin and said proteins are encoded by *tyrA* and ORF438.

18. The method of Claim 16 wherein said pigment is indigo and said proteins are *tna* and *pig*.

19. The method of Claim 16 wherein said pigment is anythocyanin and said constructs comprise the anthocyanin R and C1 regulatory genes.

20. The method of Claim 13 wherein plant cell is a cotton fiber cell, and wherein said transcriptional region is a fiber tissue transcription iniation region.

21. The method of Claim 20 wherein said transcriptional region is selected from the group consisting of the Ltp, the 4-4 and the *rac* promoter sequences

22. A recombinant DNA construct comprising the cotton tissue transcriptional sequence shown in Figure 2.

23. A recombinant DNA construct comprising the cotton tissue transcriptional sequence shown in Figure 5.

24. An isolated DNA encoding sequence of Figure 1.

25. An isolated DNA encoding sequence of Figure 4.

26. The method of Claim 13 wherein said protein of interest is involved in the synthesis of a plant hormone.

27. An isolated DNA sequence comprising the cotton lipid transfer protein encoding sequence of Figure 7.

28. A cotton fiber cell comprising a DNA sequence, wherein said DNA sequence comprises as operably joined components in the direction of transcription, a cotton fiber transcriptional factor and an open reading frame encoding a protein required for synthesis of a pigment.

29. A cotton fiber cell according to Claim 27 comprising pigment produced by said pigment synthesizing protein.

30. A cotton fiber cell according to Claim 27 wherein said DNA sequence further comprises a transport signal encoding a sequence from a plant nuclear-encoded gene.

31. A cotton fiber cell according to Claim 29, wherein said transport signal encoding sequence comprises a plastid transit peptid.

32. A cotton fiber cell according to Claim 29, wherein said transport signal encoding sequence encodes a signal peptide which provides for transport across the rough endoplasmic reticulum.

33. A cotton fiber cell according to Claim 31, wherein said sequence further comprises, 3' to said open reading frame, a vacuolar localization signal.

34. A cotton fiber cell according to Claim 27 wherein said transcriptional factor is selected from the group consisting of the cotton fiber lipid transfer promoter sequence, the 4-4 promoter sequence and the *rac* promoter sequence.

35. A cotton fiber cell according to Claim 27 wherein said pigment is melanin or indigo.

36. A cotton fiber cell according to Claim 27 wherein said open reading frame is from a bacterial gene.

37. A cotton fiber cell according to Claim 35 wherein said bacterial gene is selected from the group consisting of ORF438, *tyr*A, anthocyanin R gene, anthocyanin C1 gene, *pig*, and *tna*.

38. A cotton fiber cell comprising melanin.

39. A cotton fiber cell comprising indigo.

40. A cotton fiber cell which is colored by genetic engineering and which has a negative a* value less than - 1.0 as measured on the L*a*b* color space.

41. The cotton fiber cell of Claim 39 wherein said negative a* value is less than a -5.0.

42. The cotton fiber cell of Claim 40 wherein said negative a* value is less than a -8.0.

43. A cotton fiber cell which is colored by genetic engineering and which has an a* value less than 2 and the b* value greater than 10 as measured on the L*a*b* color space.

44. A cotton fiber cell which is colored by genetic engineering and which has a hue angle value h of greater than 100° as measured on the L*C*h color space.

45. The cotton fiber cell of Claim 43 wherein said h value is greater than a 135°.

```
                          20                                      40
CTT TCT ATT TGG TTA ACC ATG GCT CAT AAC TTT CGT CAT CCT TTC TTC
Leu Ser Ile Trp Leu Thr Met Ala His Asn Phe Arg His Pro Phe Phe>

                 60                                80
CTT TTC CAA CTT TTA CTC ATT ACT GTC TCA CTA ATG ATC GGT AGC CAC
Leu Phe Gln Leu Leu Leu Ile Thr Val Ser Leu Met Ile Gly Ser His>

   100                                120                                     140
    *
ACC GTC TCG TCA GCG GCT CGA CAT TTA TTC CAC ACA CAA ACA ACC TCA
Thr Val Ser Ser Ala Ala Arg His Leu Phe His Thr Gln Thr Thr Ser>

                         160                           180
TCA GAG CTG CCA CAA TTG GCT TCA AAA TAC GAA AAG CAC GAA GAG TCT
Ser Glu Leu Pro Gln Leu Ala Ser Lys Tyr Glu Lys His Glu Glu Ser>

        200                                   220                                    240
         *
GAA TAC AAA CAG CCA AAA TAT CAT GAA GAG TAC CCA AAA CAT GAG AAG
Glu Tyr Lys Gln Pro Lys Tyr His Glu Glu Tyr Pro Lys His Glu Lys>

                            260                                 280
CCT GAA ATG TAC AAG GAG GAA AAA CAA AAA CCC TGC AAA CAT CAT GAA
Pro Glu Met Tyr Lys Glu Glu Lys Gln Lys Pro Cys Lys His His Glu>

                  300                                 320
                   *
GAG TAC CAC GAG TCA CGC GAA TCG AAG GAG CAC GAA GAG TAC GAT AAA
Glu Tyr His Glu Ser Arg Glu Ser Lys Glu His Glu Glu Tyr Asp Lys>

  340                                 360                                380
GAA AAA CCC GAT TTC CCC AAA TGG GAA AAG CCT AAA GAG CAC GAG AAA
Glu Lys Pro Asp Phe Pro Lys Trp Glu Lys Pro Lys Glu His Glu Lys>

                      400                               420
```

FIGURE 1A

```
                         *
CAC GAA GTC GAA TAT CCG AAA ATA CCC GAG TAC AAG GAC AAA CAA GAT
His Glu Val Glu Tyr Pro Lys Ile Pro Glu Tyr Lys Asp Lys Gln Asp>

      440                         460                         480
GAG AAT AAG AAA CAT AAA GAT GAA GAG TGC CAG GAG TCA CAC GAA TCG
Glu Asn Lys Lys His Lys Asp Glu Glu Cys Gln Glu Ser His Glu Ser>

                          500                         520
                            *
AAA GAG CAC GAA GAG TAC GAG AAA GAA AAA CCC GAT TTC CCC AAA TGG
Lys Glu His Glu Glu Tyr Glu Lys Glu Lys Pro Asp Phe Pro Lys Trp>

            540                         560
GAA AAG CCT AAA GGG CAC GAG AAA CAT AAA GCC GAA TAT CCG AAA ATA
Glu Lys Pro Lys Gly His Glu Lys His Lys Ala Glu Tyr Pro Lys Ile>

  580                         600                         620
                                *
CCT GAG TGC AAG GAA AAA CTA GAT GAG GAT AAG GAA CAT AAA CAT GAG
Pro Glu Cys Lys Glu Lys Leu Asp Glu Asp Lys Glu His Lys His Glu>

                  640                         660
TTC CCA AAG CAT GAA AAA GAA GAG GAG AAG AAA CCT GAG AAA GGC ATA
Phe Pro Lys His Glu Lys Glu Glu Glu Lys Lys Pro Glu Lys Gly Ile>

      680                         700                         720
                                    *
GTA CCC TGA GTG GGT TAA AAT GCC TGA ATG GCC GAA GTC CAT GTT TAC
Val Pro *** Val Gly *** Asn Ala *** Met Ala Glu Val His Val Tyr>

                          740                         760
TCA GTC TGG CTC GAG CAC TAA GCC TTA AGC CAT ATG ACA CTG GTG CAT
Ser Val Trp Leu Glu His *** Ala Leu Ser His Met Thr Leu Val His>

            780                         800
                                          *
```

FIGURE 1B

```
GTG CCA TCA TCA TGC AGT AAT TTC ATG GGA TAT TGT AAT TAT ATT GTT
Val Pro Ser Ser Cys Ser Asn Phe Met Gly Tyr Cys Asn Tyr Ile Val>

  820                            840                           860
AAT AAA AAA GAT GGT GAG TGG GAA ATG TGT GTG TGC ATT CAT CCA TGA
Asn Lys Lys Asp Gly Glu Trp Glu Met Cys Val Cys Ile His Pro ***>

                    880                        900
                                                *
GCA ATG CTG AAT CTC TTT GCA TGC ATA GAG ATT CTG AAT GGT TAT AGT
Ala Met Leu Asn Leu Phe Ala Cys Ile Glu Ile Leu Asn Gly Tyr Ser>

        920                           940                       960
TTA TGT TAT ATC GTT TGT TCT AGT GAA ATT AAT TTT GAA TGT TGT ATG
Leu Cys Tyr Ile Val Cys Ser Ser Glu Ile Asn Phe Glu Cys Cys Met>

TAA TGT T
*** Cys Xxx>
```

FIGURE 1C

```
                            20                     40                     60
ACTAAAGGGA ACAAAAGCTG GAGCTCCACC GCGGTGGCGG CCGCTCTAGA ACTAGTGGAT

                            80                    100                    120
                                                    *
CCCCCGTGGA CTAAACAAAA CATGGGAAGA TTTGCTGTAA AAAAATAAAA GAAGCTTACT

                           140                    160                    180
CAATAACACT TTGTGAATTG TATACAAAAG ACTCAATGAA AAACAATAAC TCAATACACT

                           200                    220                    240
                             *
TTTTTTCACT GATTTACATC CTTTATATAG GCTGAAACTA CAACAACTTT AGCTAAAAAA

                           260                    280                    300
                                                                           *
ATAGGATAAC CTAATAGCAA AATCACAATC AGATATTAAA CCATGATTTT AGCTAACCAT

                           320                    340                    360
TTAACAACTT TATTGAAACT AATTTGAATA TTTCATCTGC TGATATGCCC AAGATTTTAG

                           380                    400                    420
                                                    *
GCCACTAACC GATTTGGTGG TGAACTTTAA CATGTCATGC ATTTGTAACT GTTTGAAACA

                           440                    460                    480
AGTTTTTTGC ATTATTTTAC TATATGAACT GTTTGATTAG GTTGAGTTAC ACACTGAGCT

                           500                    520                    540
                             *
TGTAAGCTCA CTCAAATTTT TCTAATTTCT AAGGTGATCA GCAAACTTAG GACCGGGCGG

                           560                    580                    600
                                                                           *
CGTACGAGAG CTCGGATTGA TTTTCTAGTT AATAAATAAG ACGATTTATG TTTTTAAACT
```

Figure 2A

```
                    620                               640                               660
ATTATGGACT TTTTGGACTA TGTAACTGTT TGGGACTTTA TTTTTGTTTT TTATTTGCTT
                    680                               700                               720
                                                         *
TTTTTGGATT TAGTAATTAT TATTTTTAAA CTGCAAAATT ATATGTTTTT ACAAACTAAG
                    740                               760                               780
TCACAGTTTT CAAAATTCCA TAACTTAGAA TTTTTCGCTG CAAAATAAAG TAATCATTTA
                    800                               820                               840
                     *
AGTGTTTTTT CTGTAATAAA ATAAATAAAT AATTTTAACG AGTATTTTCC TAAAAATTGG
                    860                               880                               900
                                                                                          *
AAATTGATTT ACCAAAATTA GTATGTCAAA ACACATGTTT ATATGTTACA GGGCGATATC
                    920                               940                               960
GTCTAGGCAA ATAACATCTA GGCGGGGTTT GGAGTGTTAC AGGGCGAGTG GGCTCATTTT
                    980                              1000                              1020
                                                         *
GAGTAAGTAT AGTTAGGGCC GAGTTTTAGA TTGCATATTC AAGGTCAAAG ATTTTGTAAA
                   1040                              1060                              1080
CTTCGATGAA TGATATGTAT GATTGTCCGA TTAACGAAAT ATGTTTTTTT CTTTTGTGTG
                   1100                              1120                              1140
                     *
TGTTTTATCT CGTGTGATAA GTATATAGTA TGTTTTATTC CAATTCTTAT GGCATGTGAC
                   1160                              1180                              1200
                                                                                          *
ATTGTGGCTA TTCTAATTAA ATTGATTTGT TATTATTGAA ATCTGATGCA TCTGTTCTAC
                   1220                              1240                              1260
```

Figure 2B

```
AAAGCATGGA ATCTCATGCC TACTGCTTTC TGTTAAAGAT ACGATTGCAA GTTTAACATG

                 1280                1300                1320
                                        *
CTTACTATTT TGATTTTGTC CTTGCATGCT ATGTCACATT ACATGGGGTT GGGATGATAT

                 1340                1360                1380
GGTAAGGAGG AAGTTTTGAC AGTTTAATGA TTTGCACTAT CTGGTGGTTT AACCACATAT

                 1400                1420                1440
                    *
TTGTTATGGC ATCTTGACTG CGGTTATGGT GGCTCGACCG CCCATATCTG TTCTGGAAAT

                 1460                1480                1500
                                                            *
TTATCTGTGA CTCTGGTGGC ATTGTCTACA ATTATTTGTT GGTGTGTTTT GGATGGACGA

                 1520                1540                1560
GTCGTGGGGA ACTCTATTTG GTGTGTTGCG GAGTTGGGTA GGAAATTTTC GAAAAAAATT

                 1580                1600                1620
                                        *
TGCATTGTGT TTTTCTGAAA AATATTGCAT TAACATAATC ATGCATTCTC AATTTTGGTC

                 1640                1660                1680
AATTGAACGT TATAAAATTC TCTATGATAT CCTGATCTGT TTATTACATT ATATGTGTTT

                 1700                1720                1740
                    *
ATGCTTGAGT TAAGTCAAAC ATTGAGATTC ATAGCTCACC CAATTATTTA ATCATTTCAG

                 1760                1780                1800
                                                            *
GCAATCTGCA GACTTAGGAT TGGATGGCGT TCAGGAGCTT GGATTGGTTT TCTCACATCA

                 1820                1840                1860
TATTTTATTA AATAATTATT AATTAAAATT TATGGACTTT TGGACTGTCT GACTAATTTT
```

Figure 2C

```
                        1880                 1900                 1920
                                                *
CAGAATTTTA TTTTGGTTTT GGGTTTTGTT GAATTTTTTA GATAATTATT TTAAATATTC

                        1940                 1960                 1980
TGCATAATTT TTCTGTTATT TGAAAAGGAT GTTCGAATTT TTTTTCAAAA TTGAAACGTT

                        2000                 2020                 2040
                           *
TAAGAATTTT TACTACTGCA AATTCAGAAT AAGTGAATTT GTTTTTTAGA AAGATTAAAT

                        2060                 2080                 2100
                                                                     *
AAGTTAGTAT TACGATTTTT AGTTTGATTT GGTGGAAAGT AATGTATGTT TTTGAACATA

                        2120                 2140                 2160
ATTATTTGAC AATAATTAAG TTTTCTAGGG AATAAACGGA AATATCTTCT TCTTTTTTGT

                        2180                 2200                 2220
                                                *
AAAATTACTA ATGCAAGAAC AAACAACGTT TTGGGGAGCA AATAATCTAG CTTTAAGTAG

                        2240                 2260                 2280
TCAGTGTAAC TCTCAAAATC TGGTCATAAC TTCTAGGCTG AGTTTGCTGT GCTACAGTAG

                        2300                 2320                 2340
                           *
TAAGTCTATA GAAACTTACC TGACAAAACG ACATGACGTC AGGGTCGAAT CTACAACTTT

                        2360                 2380                 2400
                                                                     *
TCCTTTTTCT TCAATTAACA TATGGTTGAT TCAAGTTCCG ATCTATAATA ATTTATTACG

                        2420                 2440                 2460
ATTTATCAAT TTCAATTACC TTATATCATC CTATTATAAA TATAAGTCAG TTCAATTCAG
```

Figure 2D

```
                    2480                 2500                 2520
                                            *
TTTTCGAAAG TTCCCAAAAA TTTTGAATTT TATTAAATTT ATTCCCTAAA ACCGAAATAG

                    2540                 2560                 2580
TTATATCTTT CAAATTTAAG TTTCATTTTT CAATCCGATT TCAATTTCAT CCTTTTATAA

                    2600                 2620                 2640
                       *
CTCTCTATTA TCTATAATTA CATAAATTTC AAATTAATTT TGAAATATTT ACACTTTAGT

                    2660                 2680                 2700
                                                                 *
CCCTAAGTTC AAAACTATAA ATTTTCACTT TAGAAATTAA TCATTTTTCA CATCTAAGCA

                    2720                 2740                 2760
TCAAATTTAA CCAAATGACA CAAATTTCAT GATTAGTTAG ATCAAGCTTT TGAGTCTTCA

                    2780                 2800                 2820
                                            *
AAACATAAAA ATTACAAAAA AAAAACAAAC TTAAAATCAT TTATCAATTT GAACAACAAA

                    2840                 2860                 2880
GCTTGGCCGA ATGCTAAGAG CTTAAAAATG GCTTCTTTTG TTTCTTTTTG TTGCAAACGG

                    2900                 2920                 2940
                       *
TGGAGAGAAG AGGGAAATGA AGATTGACCA TATTTTTTTA TTATGTTTTA ACATATAATA

                    2960                 2980                 3000
                                                                 *
TTAATAATTT AATCATAATT ATACTTTGGT GAATGTGACA GTGGGGAGAT ACGTAAAGTA

                    3020                 3040                 3060
TTTTAACATT ATACTTTTTG CAAGCAGTTG GCTGGTCTAC CCAAGAGTGA TCAAAGTTTG

                    3080                 3100                 3120
```

Figur 2E

```
                                           *
AGCTGCCTTC AATGAGCCAA TTTTTGCCCA TAATGGATAA AGGCAATTTG TTTAGTTCAA

                 3140                  3160                  3180
CTGCTCACAG AATAATGTTA AAATGAAATT AAAATAAGGT GGCCTGGTCA CACACACAAA

                 3200                  3220                  3240
                    *
AAAAAACTAA TGTTGGTTGG TTGAATTTTA TATTACGGAA TGTAATATTA TATTTTAAAA

                 3260                  3280                  3300
                                                                *
TAAAATTATG TTATTTAGAT TCTTAATATT TTGGAGCATT CCATACTATA ATTTCGTAAC

                 3320                  3340                  3360
ATAATATTAA AATATAGTAA TATAAAGTGT AATTAACTTT AAATTACAAG CATAATATTA

                 3380                  3400                  3420
                                          *
AATTTTGAAT CAATTAATTT TTATTTCTAT TATTTTAATT AATTTAGTCT ATTTTTTCAA

                 3440                  3460                  3480
AATAAAATTT AAATCTAAAT AAAAATAATT TTTCCTTAAT GTTGAAACAA CTCATGTTAT

                 3500                  3520                  3540
                    *
ACTTCAAAAT TATAAGTATT ATATTTACCT TGATGATTTA TTTATTAGTA TATTAATTCT

                 3560                  3580                  3600
                                                                *
GATTATAATT ATGGTGGGAT ACAATCGCTT TCCACTAAAT ATTTTAACTA TGATTTATAA

                 3620                  3640                  3660
ATTTATTTCA ACATCGTATA TTTACTTATT AATACATAAT TTATCATAAT TTTATGGAAA

                 3680                  3700                  3720
                                          *
```

Figure 2F

```
TTGAGACCAA GAAACATTAA GAGAACAAAT TCTATAACAA AGACAATTTA GAAAAAAATG

                 3740                 3760                 3780
TACTTTTAGG TAATTTTAAG TACTCTTAAC CAAACACAAA AATTCAAATC AAATGAACTA

                 3800                 3820                 3840
                    *
AATAAGATAA TATAACATAC GGAACATCTT ACTTGTAATC TTACATTCCC ATAATTTTAT

                 3860                 3880                 3900
                                                              *
TATGAAAAAT AATCTTATAT TACTCGAACT AAATGTTGTC ACAAATTATT ATCTAAATAA

                 3920                 3940                 3960
AGAAAAACAC TTAATTTTTA TAACATTTTT TCATATATTT GAAAGATTAT ATTTTGTATA

                 3980                 4000                 4020
                                         *
TTTACGTAAA AATATTTGAC ATAGATTGAG CACCTTCTTA ACATAATCCC ACCATAAGTC

                 4040                 4060                 4080
AAGTATGTAG ATGAGAAATT GGTACAAACA ACGTGGGGCC AAATCCCACC AAACCATCTC

                 4100                 4120
                    *
TCATTCTCTC CTATAAAAGG CTTGCTACAC ATAGACAACA ATCCACACA C AAA TAC
                                                       <Phe Val

 4140                    4160                    4180
ACG TTC TTT TCT TTC TAT TTG ATT AAC CAT GGC TCA TAG CAT TCG TCA
<Arg Glu Lys Arg Glu Ile Gln Asn Val Met Ala *** Leu Met Arg ***

                4200                    4220
                   *
CCC TTT CTT CCT TTT CCA ACT TTT ACT CAT AAG TGT CTC ACT AGT GAC
<Gly Lys Lys Arg Lys Trp Ser Lys Ser Met Leu Thr Glu Ser Thr Val
```

Figure 2G

```
        4240                              4260                          4280
CGG TAG CCA CAC TGT TTC GGC AGC GGC TCG ACG TTT ATT CGA GAC ACA
<Pro Leu Trp Val Thr Glu Ala Ala Ala Arg Arg Lys Asn Ser Val Cys

                            4300                          4320
                              *
AGC AAC CTC ATC AGA GCT CCC ACA ATT GGC TTC AAA ATA CGA AAG CAC
<Ala Val Glu Asp Ser Ser Gly Cys Asn Ala Glu Phe Tyr Ser Leu Val

              4340                          4360
GAG AGT CTG AAT ACG AAA AGC CAG AAT ACA AAC AGC CAA AGT ATC ACG
<Leu Thr Gln Ile Arg Phe Ala Leu Ile Cys Val Ala Leu Thr Asp Arg

  4380                          4400                          4420
                                  *
AAG AGT ACT CAA AAC TTG AGA AGC CTG AAA TGC AAA AGG AGG AAA AAC
<Leu Thr Ser Leu Val Gln Ser Ala Gln Phe Ala Phe Pro Pro Phe Val

                    4440                          4460
AAA AAC CCT GCA AAC AGC ATG AAG AGT ACC ACG AGT CAC ACG AAT CAA
<Phe Val Arg Cys Val Ala His Leu Thr Gly Arg Thr Val Arg Ile Leu

        4480                          4500                          4520
                                        *
AGG AGC AAA AAG AGT ACG AGA AAG AAA ATC TCGACGAA TTCCCCCGGG
<Pro Ala Phe Leu Thr Arg Ser Leu Phe Asp

                    4540                    4560                    4580
CGTCGACGGC TAGCGAAGAT CTTCGGGCCC GTCGAGCCTT GAATCATATG ACACTGGTGC

                    4600                    4620                    4640
                       *
ATGTGCCATC ATCATGCAGT AATTTCATGG TATATCGTAA TATATAGTTA ATAAAAAAGA

                    4660                    4680                    4700
                                                                       *
TGGTGATTGG GAAATGTGTG TGTGCATTCC TCCATGCACT AATGGTGAAT CTCTTTGCAT
```

Figure 2H

```
                              4720                           4740                           4760
ACATAGAAAT TCTAAATGGT TATAGTTTAT GTTATAGTGT ATGTTGTAGT GAAATTAATT

                              4780                           4800                           4820
                                                                 *
TTAAATGTTG TATCTAATGT TAACATCACT TGGCTTGATT TATGTTATGT TATGTATTTT

                              4840                           4860                           4880
ACTTTAATGA TATTGCATGT ATTGTTAATT TAACATTGCT TGATCATTAT ACTCTTCTAC

                              4900                           4920                           4940
                                 *
TATTAATTAT AAATGGCACT GTTTTGTTTA AACTTTTTAC AAGTTAAGAC ATGTATAAAT

                              4960                           4980                           5000
                                                                                               *
ATATGACAAT ATAATTACAG GTTTTAGTTC AATGTTAGCT ATCTTAGTAT GTTATTGATG

                              5020                           5040                           5060
ATCTTAATTA CATTTAAACA AATTCCACTT AAAATTTTAA TAAATAATAA CAAATAATTA

                              5080                           5100                           5120
                                                                 *
TTGTAATATA ATACATTAAA TGCAACAAAA AATGAAATAA ATAAAATAAA ATAGCAAATA

                              5140                           5160                           5180
ATTGTTATAA TATTGTAATA TAATATGTAC CATATTCTTA ACTGAAATAG GGTCTAACCT

                              5200                           5220                           5240
                                 *
ATAATCCCTA AAATTTCAGT TTAAATATTT TTATACCTAC CATATTATTA GAACTCTTTT

                              5260                           5280                           5300
                                                                                               *
TAAATATATT AAAATTTTAA TTATACCAAT TTAATTAAAC TATTAATTAT CTTAACTAAA
```

Figure 2I

```
                     5320                5340                5360
ATCTAAAATT TTATTTAACC TATTAATAAA TTCCTAATTA TCTTATCTAA TTTAAAACTC

                     5380                5400                5420
                                            *
TAATTATCCT AATTTAATTT AAATTCTTAA TTATCTTAAT TTGTAACCTC CTCCACCCAG

                     5440                5460                5480
CTAGATGCTG GACCCGAATC CGGGAGATTA CATCGGCCAT TGAGATGGCG TGATCAGGGT

                     5500                5520                5540
                        *
TTGGCGCGCC GGTACCCAAT TCGCCCTATA GTGAGTTCGT ATTACGCGCG CTCACTGCGT

CCGGTTT
```

Figure 2J

```
                    20                  40                  60
ACTAAAGGGA ACAAAAGCTG GAGCTCCACC GCGGTGGCGG CCGCTCTAGG ATCCCCCGTG

                    80                 100                 120
                                         *
GACTAAACAA AACATGGGAA GATTTGCTGT AAAAAAATAA AAGAAGCTTA CTCAATAACA

                   140                 160                 180
CTTTGTGAAT TGTATACAAA AGACTCAATG AAAAACAATA ACTCAATACA CTTTTTTTCA

                   200                 220                 240
                     *
CTGATTTACA TCCTTTATAT AGGCTGAAAC TACAACAACT TTAGCTAAAA AAATAGGATA

                   260                 280                 300
                                                             *
ACCTAATAGC AAAATCACAA TCAGATATTA AACCATGATT TTAGCTAACC ATTTAACAAC

                   320                 340                 360
TTTATTGAAA CTAATTTGAA TATTTCATCT GCTGATATGC CCAAGATTTT AGGCCACTAA

                   380                 400                 420
                                         *
CCGATTTGGT GGTGAACTTT AACATGTCAT GCATTTGTAA CTGTTTGAAA CAAGTTTTTT

                   440                 460                 480
GCATTATTTT ACTATATGAA CTGTTTGATT AGGTTGAGTT ACACACTGAG CTTGTAAGCT

                   500                 520                 540
                     *
CACTCAAATT TTTCTAATTT CTAAGGTGAT CAGCAAACTT AGGACCGGGC GGCGTACGAG

                   560                 580                 600
                                                             *
AGCTCGGATT GATTTTCTAG TTAATAAATA AGACGATTTA TGTTTTTAAA CTATTATGGA
```

Figure 3A

```
                       620                    640                    660
CTTTTTGGAC TATGTAACTG TTTGGGACTT TATTTTTGTT TTTTATTTGC TTTTTTTGGA

                       680                    700                    720
                                                *
TTTAGTAATT ATTATTTTTA AACTGCAAAA TTATATGTTT TTACAAACTA AGTCACAGTT

                       740                    760                    780
TTCAAAATTC CATAACTTAG AATTTTTCGC TGCAAAATAA AGTAATCATT TAAGTGTTTT

                       800                    820                    840
                         *
TTCTGTAATA AAATAAATAA ATAATTTTAA CGAGTATTTT CCTAAAAATT GGAAATTGAT

                       860                    880                    900
                                                                       *
TTACCAAAAT TAGTATGTCA AAACACATGT TTATATGTTA CAGGGCGATA TCGTCTAGGC

                       920                    940                    960
AAATAACATC TAGGCGGGGT TTGGAGTGTT ACAGGGCGAG TGGGCTCATT TTGAGTAAGT

                       980                   1000                   1020
                                                *
ATAGTTAGGG CCGAGTTTTA GATTGCATAT TCAAGGTCAA AGATTTTGTA AACTTCGATG

                      1040                   1060                   1080
AATGATATGT ATGATTGTCC GATTAACGAA ATATGTTTTT TTCTTTTGTG TGTGTTTTAT

                      1100                   1120                   1140
                         *
CTCGTGTGAT AAGTATATAG TATGTTTTAT TCCAATTCTT ATGGCATGTG ACATTGTGGC

                      1160                   1180                   1200
                                                                       *
TATTCTAATT AAATTGATTT GTTATTATTG AAATCTGATG CATCTGTTCT ACAAAGCATG

                      1220                   1240                   1260
```

Figure 3B

```
GAATCTCATG CCTACTGCTT TCTGTTAAAG ATACGATTGC AAGTTTAACA TGCTTACTAT
                1280                   1300                   1320
                                          *
TTTGATTTTG TCCTTGCATG CTATGTCACA TTACATGGGG TTGGGATGAT ATGGTAAGGA
                1340                   1360                   1380
GGAAGTTTTG ACAGTTTAAT GATTTGCACT ATCTGGTGGT TTAACCACAT ATTTGTTATG
                1400                   1420                   1440
                   *
GCATCTTGAC TGCGGTTATG GTGGCTCGAC CGCCCATATC TGTTCTGGAA ATTTATCTGT
                1460                   1480                   1500
                                                                 *
GACTCTGGTG GCATTGTCTA CAATTATTTG TTGGTGTGTT TTGGATGGAC GAGTCGTGGG
                1520                   1540                   1560
GAACTCTATT TGGTGTGTTG CGGAGTTGGG TAGGAAATTT TCGAAAAAAA TTTGCATTGT
                1580                   1600                   1620
                                          *
GTTTTTCTGA AAAATATTGC ATTAACATAA TCATGCATTC TCAATTTTGG TCAATTGAAC
                1640                   1660                   1680
GTTATAAAAT TCTCTATGAT ATCCTGATCT GTTTATTACA TTATATGTGT TTATGCTTGA
                1700                   1720                   1740
                   *
GTTAAGTCAA ACATTGAGAT TCATAGCTCA CCCAATTATT TAATCATTTC AGGCAATCTG
                1760                   1780                   1800
                                                                 *
CAGACTTAGG ATTGGATGGC GTTCAGGAGC TTGGATTGGT TTTCTCACAT CATATTTTAT
                1820                   1840                   1860
TAAATAATTA TTAATTAAAA TTTATGGACT TTTGGACTGT CTGACTAATT TTCAGAATTT
```

Figure 3C

```
                    1880                1900                1920
                                           *
TATTTTGGTT TTGGGTTTTG TTGAATTTTT TAGATAATTA TTTTAAATAT TCTGCATAAT

                    1940                1960                1980
TTTTCTGTTA TTTGAAAAGG ATGTTCGAAT TTTTTTTCAA AATTGAAACG TTTAAGAATT

                    2000                2020                2040
                       *
TTTACTACTG CAAATTCAGA ATAAGTGAAT TTGTTTTTTA GAAAGATTAA ATAAGTTAGT

                    2060                2080                2100
                                                               *
ATTACGATTT TTAGTTTGAT TTGGTGGAAA GTAATGTATG TTTTTGAACA TAATTATTTG

                    2120                2140                2160
ACAATAATTA AGTTTTCTAG GGAATAAACG GAAATATCTT CTTCTTTTTT GTAAAATTAC

                    2180                2200                2220
                                           *
TAATGCAAGA ACAAACAACG TTTTGGGGAG CAAATAATCT AGCTTTAAGT AGTCAGTGTA

                    2240                2260                2280
ACTCTCAAAA TCTGGTCATA ACTTCTAGGC TGAGTTTGCT GTGCTACAGT AGTAAGTCTA

                    2300                2320                2340
                       *
TAGAAACTTA CCTGACAAAA CGACATGACG TCAGGGTCGA ATCTACAACT TTTCCTTTTT

                    2360                2380                2400
                                                               *
CTTCAATTAA CATATGGTTG ATTCAAGTTC CGATCTATAA TAATTTATTA CGATTTATCA

                    2420                2440                2460
ATTTCAATTA CCTTATATCA TCCTATTATA AATATAAGTC AGTTCAATTC AGTTTTCGAA
```

Figure 3D

```
                    2480                2500                2520
                                           *
AGTTCCCAAA AATTTTGAAT TTTATTAAAT TTATTCCCTA AAACCGAAAT AGTTATATCT
                    2540                2560                2580
TTCAAATTTA AGTTTCATTT TTCAATCCGA TTTCAATTTC ATCCTTTTAT AACTCTCTAT
                    2600                2620                2640
                       *
TATCTATAAT TACATAAATT TCAAATTAAT TTTGAAATAT TTACACTTTA GTCCCTAAGT
                    2660                2680                2700
                                                               *
TCAAAACTAT AAATTTTCAC TTTAGAAATT AATCATTTTT CACATCTAAG CATCAAATTT
                    2720                2740                2760
AACCAAATGA CACAAATTTC ATGATTAGTT AGATCAAGCT TTTGAGTCTT CAAAACATAA
                    2780                2800                2820
                                           *
AAATTACAAA AAAAAAACAA ACTTAAAATC ATTTATCAAT TTGAACAACA AAGCTTGGCC
                    2840                2860                2880
GAATGCTAAG AGCTTAAAAA TGGCTTCTTT TGTTTCTTTT TGTTGCAAAC GGTGGAGAGA
                    2900                2920                2940
                       *
AGAGGGAAAT GAAGATTGAC CATATTTTTT TATTATGTTT TAACATATAA TATTAATAAT
                    2960                2980                3000
                                                               *
TTAATCATAA TTATACTTTG GTGAATGTGA CAGTGGGGAG ATACGTAAAG TATTTTAACA
                    3020                3040                3060
TTATACTTTT TGCAAGCAGT TGGCTGGTCT ACCCAAGAGT GATCAAAGTT TGAGCTGCCT
                    3080                3100                3120
```

Figure 3E

```
                                   *
TCAATGAGCC AATTTTTGCC CATAATGGAT AAAGGCAATT TGTTTAGTTC AACTGCTCAC
                 3140                  3160                  3180
AGAATAATGT TAAAATGAAA TTAAAATAAG GTGGCCTGGT CACACACACA AAAAAAAACT
                 3200                  3220                  3240
                    *
AATGTTGGTT GGTTGAATTT TATATTACGG AATGTAATAT TATATTTTAA AATAAAATTA
                 3260                  3280                  3300
                                                                *
TGTTATTTAG ATTCTTAATA TTTTGGAGCA TTCCATACTA TAATTTCGTA ACATAATATT
                 3320                  3340                  3360
AAAATATAGT AATATAAAGT GTAATTAACT TTAAATTACA AGCATAATAT TAAATTTTGA
                 3380                  3400                  3420
                                          *
ATCAATTAAT TTTTATTTCT ATTATTTTAA TTAATTTAGT CTATTTTTTC AAAATAAAAT
                 3440                  3460                  3480
TTAAATCTAA ATAAAAATAA TTTTTCCTTA ATGTTGAAAC AACTCATGTT ATACTTCAAA
                 3500                  3520                  3540
                    *
ATTATAAGTA TTATATTTAC CTTGATGATT TATTTATTAG TATATTAATT CTGATTATAA
                 3560                  3580                  3600
                                                                *
TTATGGTGGG ATACAATCGC TTTCCACTAA ATATTTTAAC TATGATTTAT AAATTTATTT
                 3620                  3640                  3660
CAACATCGTA TATTTACTTA TTAATACATA ATTTATCATA ATTTTATGGA AATTGAGACC
                 3680                  3700                  3720
                                          *
```

Figure 3F

```
AAGAAACATT AAGAGAACAA ATTCTATAAC AAAGACAATT TAGAAAAAAA TGTACTTTTA

                 3740                3760                3780
GGTAATTTTA AGTACTCTTA ACCAAACACA AAAATTCAAA TCAAATGAAC TAAATAAGAT

                 3800                3820                3840
                    *
AATATAACAT ACGGAACATC TTACTTGTAA TCTTACATTC CCATAATTTT ATTATGAAAA

                 3860                3880                3900
                                                            *
ATAATCTTAT ATTACTCGAA CTAAATGTTG TCACAAATTA TTATCTAAAT AAAGAAAAAC

                 3920                3940                3960
ACTTAATTTT TATAACATTT TTTCATATAT TTGAAAGATT ATATTTTGTA TATTTACGTA

                 3980                4000                4020
                                        *
AAAATATTTG ACATAGATTG AGCACCTTCT TAACATAATC CCACCATAAG TCAAGTATGT

                 4040                4060                4080
AGATGAGAAA TTGGTACAAA CAACGTGGGG CCAAATCCCA CCAAACCATC TCTCATTCTC

                 4100                4120
                    *
TCCTATAAAA GGCTTGCTAC ACATAGACAA CAATCCACAC A CA AAT ACA CGT TCT
                                                <Ile Cys Thr Arg

  4140                      4160                  4180
TTT CTT TCT ATT TGA TTA ACC ATG G CTCATAGCAT TCGTCACCCT TTCTTCCTTT
<Lys Lys Arg Asn Ser *** Gly His

     4200                  4220                  4240
        *
TCCAACTTTT ACTCATAAGT GTCTCACTAG TGACCGGTAG CCACACTGTT TCGGCAGCGG

     4260                  4280                  4300
```

Figure 3G

```
                                                       *
CTCGACGTTT ATTCGAGACA CAAGCAACCT CATCAGAGCT CCCACAATTG GCTTCAAAAT

      4320                  4340                  4360
ACGAAAAGCA CGAAGAGTCT GAATACGAAA AGCCAGAATA CAAACAGCCA AAGTATCACG

      4380                  4400                  4420
                               *
AAGAGTACTC AAAACTTGAG AAGCCTGAAA TGCAAAAGGA GGAAAAACAA AAACCCTGCA

      4440                  4460                  4480
AACAGCATGA AGAGTACCAC GAGTCACACG AATCAAAGGA GCAAAAAGAG TACGAGAAAG

      4500                  4520                  4540
         *
AAAATCTCGA CGGGCCCGAA GATCTTCGCT AGCCGTCGAC GCCCGGGGGA ATTCGTCGAG

      4560                  4580                  4600
                                                       *
CCTTGAATCA TATGACGCTG GTGCATGTGC CATCATCATG CAGTAATTTC ATGGTATATC

      4620                  4640                  4660
GTAATATATA GTTAATAAAA AAGATGGTGA TTGGGAAATG TGTGTGTGCA TTCCTCCATG

      4680                  4700                  4720
                               *
CACTAATGGT GAATCTCTTT GCATACATAG AAATTCTAAA TGGTTATAGT TTATGTTATA

      4740                  4760                  4780
GTGTATGTTG TAGTGAAAKT AATTTTAAAT GTTGTATCTA ATGTTAACAT CACTTGGCTT

      4800                  4820                  4840
         *
GATTTATGTT ATGTTATGTA TTTTACTTTA ATGATATTGC ATGTATTGTT AATTTAACAT

      4860                  4880                  4900
                                                       *
```

Figure 3H

```
TGCTTGATCA TTATACTCTT CTACTATTAA TTATAAATGG CACTGTTTTG TTTAAACTTT

      4920                  4940                  4960
TTACAAGTTA AGACATGTAT AAATATATGA CAATATAATT ACAAGTTTTA GTTCAATGTT

      4980                  5000                  5020
                               *
AGCTATCTTA GTATGTTATT GATGATCTTA ATTACATTTA AACAAATTCC ACTTAAAATT

      5040                  5060                  5080
TTAATAAATA ATAACAAATA ATTATTGTAA TATAATACAT TAAATGCAAC AAAAAATGAA

      5100                  5120                  5140
         *
ATAAATAAAA TAAAATAGCA AATAATTGTT ATAATATTGT AATATAATAT GTACCATATT

      5160                  5180                  5200
                                                     *
CTTAACTGAA ATAGGGTCTA ACCTATAATC CCTAAAATTT CAGTTTAAAT ATTTTTATAC

      5220                  5240                  5260
CTGCCATATT ATTAGAACTC TTTTTAAATA TATTAAAATT TTAATTATAC CAATTTAATT

      5280                  5300                  5320
                               *
TAAACTATTA ATTATCTTAA CTAAAATCTA AAATTTTATT TAACCTATTA ATTAAATTCC

      5340                  5360                  5380
TAATTATCTT ATCTAATTTA AAACTCTAAT TATCCTAATT TGATTTAAAT TCTTGATTAT

      5400                  5420                  5440
         *
CTTAATTTGT AACCTCCTCC ACCCAGCTAG ATGCTGGACC CGAATCCGGG AGATTACATC

      5460                  5480                  5500
                                                     *
GGCATTGAGA TGGCCTAGTA GTGATCAGGG TTTTCTAGAG GTACCCAATT CGCCCTATAG
```

Figure 3I

TGAGTCGT

Figure 3J

```
AAAAAACA ATG AGC ACT GCA AGA TTT ATC AAG TGT GTC ACG GTC GGT GAT    50
         Met Ser Thr Ala Arg Phe Ile Lys Cys Val Thr Val Gly Asp
          1               5                  10

GGA GCT GTG GGG AAA ACT TGT ATG CTC ATT TCA TAT ACC AGC AAT ACT     98
Gly Ala Val Gly Lys Thr Cys Met Leu Ile Ser Tyr Thr Ser Asn Thr
 15                  20                  25                  30

TTC CCA ACG GAT TAT GTT CCA ACA GTA TTT GAT AAC TTT AGT GCC AAT    146
Phe Pro Thr Asp Tyr Val Pro Thr Val Phe Asp Asn Phe Ser Ala Asn
                 35                  40                  45

GTG GTG GTG GAT GGC AGC ACA GTG AAC CTT GGC CTA TGG GAC ACT GCC    194
Val Val Val Asp Gly Ser Thr Val Asn Leu Gly Leu Trp Asp Thr Ala
             50                  55                  60

GGG CAA GAA GAT TAT AAT AGG CTA AGG CCA CTG AGT TAT AGA GGA GCT    242
Gly Gln Glu Asp Tyr Asn Arg Leu Arg Pro Leu Ser Tyr Arg Gly Ala
         65                  70                  75

GAT GTG TTT TTG TTG GCC TTT TCT CTT ATA AGC AAG GCC AGT TAT GAA    290
Asp Val Phe Leu Leu Ala Phe Ser Leu Ile Ser Lys Ala Ser Tyr Glu
     80                  85                  90

AAC ATC TAC AAA AAG TGG ATC CCA GAG CTA AGA CAT TAT GCT CAT AAT    338
Asn Ile Tyr Lys Lys Trp Ile Pro Glu Leu Arg His Tyr Ala His Asn
 95                 100                 105                 110

GTA CCA GTT GTG CTT GTT GGA ACC AAA CTA GAT TTG CGA GAT GAC AAG    386
Val Pro Val Val Leu Val Gly Thr Lys Leu Asp Leu Arg Asp Asp Lys
                115                 120                 125

CAG TTC CTC ATT GAT CAC CCT GGA GCA ACA CCA ATA TCA ACA TCT CAG    434
Gln Phe Leu Ile Asp His Pro Gly Ala Thr Pro Ile Ser Thr Ser Gln
            130                 135                 140

GGA GAA GAA CTA AAG AAG ATG ATA GGA GCA GTT ACT TAT ATA GAA TGC    482
```

FIGURE 4A

```
Gly Glu Glu Leu Lys Lys Met Ile Gly Ala Val Thr Tyr Ile Glu Cys
        145                 150                 155

AGC TCC AAA ACC CAA CAG AAT GTG AAG GCT GTT TTC GAT GCT GCA ATA      530
Ser Ser Lys Thr Gln Gln Asn Val Lys Ala Val Phe Asp Ala Ala Ile
    160                 165                 170

AAA GTA GCT TTG AGG CCA CCA AAA CCA AAG AGA AAG CCT TGC AAA AGG      578
Lys Val Ala Leu Arg Pro Pro Lys Pro Lys Arg Lys Pro Cys Lys Arg
175                 180                 185                 190

AGA ACA TGT GCT TTC CTT TGAATATTGG ATCATTATTA CAGTCAAAAA             626
Arg Thr Cys Ala Phe Leu
                195

CAGTTAACAA AAGCTGTTGC AGATAAACAC TGAATCTGCT ATAGTTTGTT TTTGGTTTAC    686

ATATGTTCCA CGTGAAACTA TGAAGCATCT CTAAGAAAAC CCAAACTATC ATATCAACCC    746

ATCGATCAAT GAATCGATTT CAATTTTCGC AGTATAAGTT CCTTTTAATC CTTTCTTTTT    806

ACTTCATTTT ATAACGAATT CTATGGATAA TGTTCCCTAC AAACATGTCA TTACAATGTT    866

TAATTATAAA TTCCATTCTT CTATTTTACT AAAAAAAAAA AAAA                     910
```

FIGURE 4B

```
                             20                     40                     60
TTGGATGAGA ACCAATTTTT AATAGTAAAN CCTAACCAAT TTTTAATAAT AAAGCTGACT

                             80                    100                    120
                                                     *
CCTAGTACAA GAGCTTTTAT TCATTCTTCT ATTTTGCTTT CCTCTAGGCT TGGCAATCGA

                            140                    160                    180
GAATTTTCTT GTGTTACAAT ATAATAAATA CATCGTAGAA ATAAATTTTA TTCAAATTGA

                            200                    220                    240
                              *
AGTCTTAACC ATCTTTAATA TTTGTAGATG TAATTTAAAT GAAAGATAAA TACATATTCT

                            260                    280                    300
                                                                            *
TGGACATGTA TTTTCATCTT AATGTTTGTG GCTTTGGTGA TAGGTGTATT GATGTACGAT

                            320                    340                    360
GTCTTTTAAA TCACATATCA CATTTTGAGT TTGTATGATG ATAAGTCGAC ATAANCGAAA

                            380                    400                    420
                                                     *
TATGGTGTGA TCTTCACTTT TGAACTTTGA TAAGTCACCA AACTTTAACA AAGTTTGATT

                            440                    460                    480
GTGTACATAT ATATATATAT CTTCAAATTT TATAATAAAA ATTGTGTTTA AATAATTTAC

                            500                    520                    540
                              *
AGTTATATTA TTTTTTTATC TCTAATTTTA TTTGTCGCCA AATTTTTAGT TGATATTTTA

                            560                    580                    600
                                                                            *
ACATAAAAAA AATTGTACAC ATTTACAAGC CCATATACAA ATAATTATAT AAATATTCAT
```

FIGURE 5/A

```
                     620                  640                  660
TAAAAAATAT ATTTAAATAT AGGATATAAA TATAACTATT TTAGAATTAT TCTACTTTAA

                     680                  700                  720
                                            *
GATAACATAG GTTAAATGTA TAATTAATAA GGTTAGTTTA TTGTAAAGAT GAGTATATAT

                     740                  760                  780
GTCGTAAACA TAATCACTAA CCATTTTTAT TAACTTCTTG GTTTTGAAGT TCCAAAAAGA

                     800                  820                  840
                       *
AAATGGAAGG GAAATTTGAG AGTAAGTTCA TGTTTATATT ATACATAATG AAGTTGATGT

                     860                  880                  900
                                                                 *
TTTCTTCTTT TTAATATTTT TATACAAAAT ATTTAAATAA AATAATTAAG GATTGAATGA

                     920                  940                  960
AAAATATAAT GAAAGTCGTT TTACTAATAG TCATATTGCA TTTTGTCGCA TCTACTTAAA

                     980                 1000                 1020
                                            *
TAATAGATAA ATTAATTGTG GTACATTAGA TCAAAGAACA AACTAGATTT TGTCCCATTC

                    1040                 1060                 1080
TATTGTTAAA AGCTGGTCCG TTTACATTAA AATAAGGTAC ATGTTACATG CCACGTATAA

                    1100                 1120                 1140
                       *
CTATCTGGTT ATTCTATCAA TCACGCTAAT TTTTAACAGT AGAAATGAAT GTAATTTTTA

                    1160                 1180                 1200
                                                                 *
AATAGAAAGG GTCAAATTGT TATTTGATCT AACACGTAGG GATTAATTTA CTTATTTTCC

                    1220                 1240                 1260
```

FIGURE 5/B

```
TAAAGAAATA AGTAAAATAT AATTTGAATC TTAATACAAA AACTTTCATG ATACTTTTAT
                1280                 1300                 1320
                                        *
CATATTTTAC TTATAATTTA ATATTGTGAG AGTAACAAAR TTAAAAAACA TAGAAACACC
                1340                 1360                 1380
AAAAGTTAGT TATGGTGTGA CTCATATACA CAGTTAAAAT TTGAATAAAT TTTTTTCTTC
                1400                 1420                 1440
                   *
GTCATTAATT CCATCATGGG TTTTTTTTTT TCTAGTTAAG CCATAATTAT CAAAATAATC
                1460                 1480                 1500
                                                             *
ATCATTAATC CTATCAATAC CCCGCCCTGC CTCCCTCCCT CAATACTTAA ACCCAACTAA
                1520                 1540                 1560
CACCCAGCAC CAAACGCACT TTAATAGCCA CCTATTTCTA GCCATGTCCT TGCACTTAAA
                1580                 1600                 1620
                                        *
GAAAAGTAAA GCTAACCTGC AATCATTCCA TATCGAGGCC TCAACAGATA AAGTTGGTTG
                1640                 1660                 1680
ATGGGTTTGC ACCAAGTTGT TAAAACCCGG CCCTCAACTT CCCTTTTCTT TTCATCCTCC
                1700                 1720                 1740
                   *
CCACTCCACA CCCTCCAATT TTCTTCATAT GGTTCTATTA TAAGTTCTTT ATAATCACAG
                1760                 1780                 1800
                                                             *
AATCAAGATA AGTCCTCAGC AAACAAAAAA CCATGGCTCT CGAGCAAGAT CTGGACTAGT
                1820                 1840                 1860
CAGAGCTCTG AATATTGGAT CATTATTACA GTCAAAAACA GTTAACAAAA GCTGTTGCAG
```

FIGURE 5/C

```
      1880                1900                1920
                             *
ATAAACACTG AATCTGCTAT AGTTTGTTTT TGGTTTACAT ATGTTCCACG TGAAACTATG

      1940                1960                1980
AAGCATCTCT AAGAAAACCC AAACTATCAT ATCAACCCAT CGATCAATGA ATCGATTTCA

      2000                2020                2040
         *
ATTTTCGCAG TATAAGTTCC TTTTAATCCT TTCTTTTTAC TTCATTTTAT AACGAATTCT

      2060                2080                2100
                                                       *
ATGGATAATG TTCCCTACAA ACATGTCATT ACAATGTTTA ATTATAAATT CCATTCTTCT

      2120                2140                2160
ATTTTACTAA GATATTAGTA ACTTCAAACT GCTGATTTTT ACTAATTTAT TATTTATAAA

      2180                2200                2220
                             *
TTGTTAGAAT GATTATTTTT CAATAATTTA ACAACAATAT TTAATATTAT TATTATTATT

      2240                2260                2280
ATTTCTCAAT TTTTATTAAA CAAAAACATA AATTTTTGAC AAATTAAAAT AAATGAATTA

      2300                2320                2340
         *
ATTTCTCAAT TTTTCGTGCA ACTATTACAA AAATCCTTCA TAGTCCTAAT CTTAATTTGA

      2360                2380                2400
                                                       *
TGCAGAGGTG ATAATAATCT TAATTTGATG CAGAGGTAAT AATGGGCCGG GTTTGAGCTG

      2420                2440                2460
GACTTAAGCA TGATATTGAC GTACTTTATA TTTTTCCAAA TTCAACCCAG CTCGAAATAT
```

FIGURE 5/D

```
                          2480                          2500                          2520
                                                           *
GAGTCTAAAA TTTTGTCCAA TTTAATCCAA GCCCATTTTA AGTTCGTCCA TATTATTTTT

                          2540                          2560                          2580
TAATTTAAAA AATTTATATC ATTTTATTTT AATATTTAAT TATTTTATAT ATTTTTTATT

                          2600                          2620                          2640
                             *
TATTGAAAAT TTTTATATAG TCATCTTAAC ATTATGTTAA TGTTTATATT AGAGTAGTAT

                          2660                          2680                          2700
                                                                                         *
TATATATATT TAGTATAGGT TTATTTTGTT AATAAACTTA AAAATGGGTC TTGTGGGCTA

                          2720                          2740                          2760
GACTTGGACC TTAAATGCTC AAACTCAAAC TTAATTCATA TTTTAAACAG GCTTAATATT

                          2780                          2800                          2820
                                                           *
TTTATTTACA CTGTTTCAAA TTTTTCGGGT GAAATATCTT CGAGTCTAGA TTAATAACAC

                          2840                          2860                          2880
CACAGGTCTA ATTTGATGCT CAATGAAAAT GAAATCATAT TGAGCTTAAT TAATATTCCA

                          2900                          2920                          2940
                             *
TTCTTCTTTG CTGAAAGGAC CAAGCAATTC GAGTTACATT AAGGTTAAAG AGTATGGGAT

                          2960                          2980                          3000
                                                                                         *
CCGCCAAACC TGCCCCAATG TCTCTTCAAC CATCCAAAAA CTTGAGTCAG TATCACATAC

                          3020                          3040
ATGTACCGNT ATTTATTTAT TTATTGAAAT TGGCATTATT TCTTG
```

FIGURE 5/E

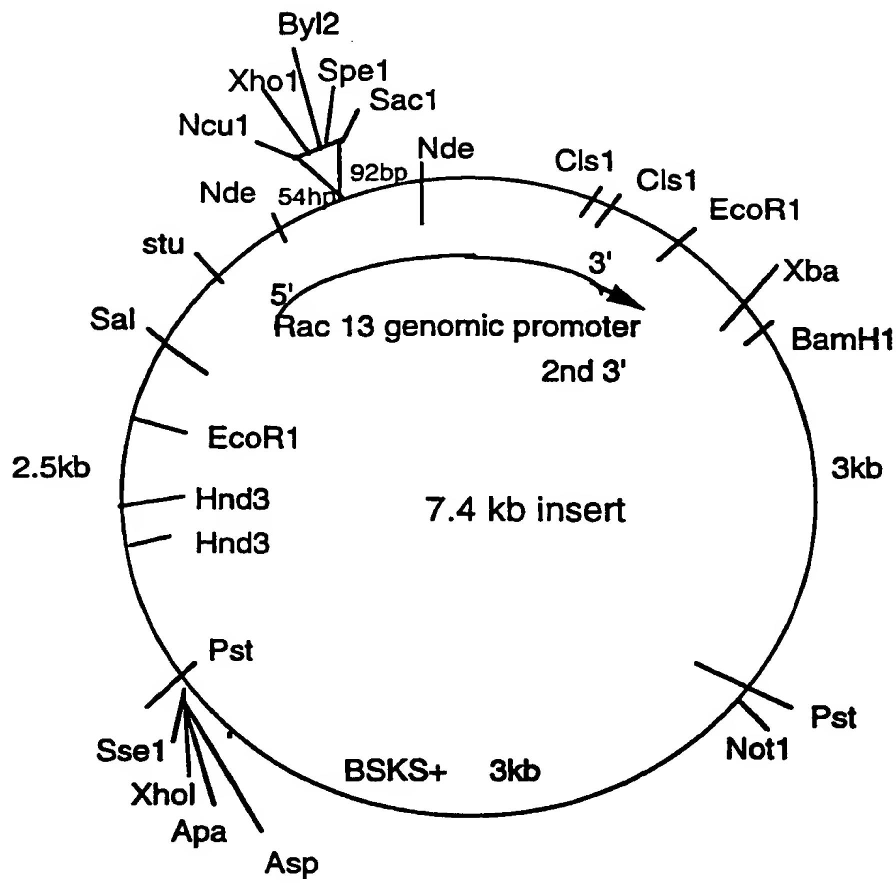
Byl2
Xho1
Spe1
Sac1
Ncu1
92bp
Nde
Cls1
Cls1
Nde
54bp
EcoR1
stu
3'
Xba
5'
Sal
Rac 13 genomic promoter
BamH1
2nd 3'
EcoR1
2.5kb
3kb
Hnd3
7.4 kb insert
Hnd3
Pst
Pst
Sse1
Not1
BSKS+
3kb
XhoI
Apa
Asp

FIGURE 6

```
GGGCATTCCA CACGACCATG TGTCCCCTAT TTCCAGGCAT TTTGAGACTT CACCTAAACT   60
TCTAGAGTTG TTTCAAATTA GCCCCTATTT GTTCTTAAAT CATTTTAGGA TCTTGTAAAC  120
TCGTATTTAG GACTAAATGT GTAATTTATA CTTTAATTAT GATTGATTAA TTGATTGATT  180
TNGTAGTAAT GCCCGTGACC CTAATCCGTT AGCGAAGAGG GGTTAGGGGT TAGGGGTTTT  240
ATTATTATTT TTTAGATATT GTATAACTCT TGTTTTATTT TTAATTTTGT TACTATTTCA  300
AAGGCATTTG TTTGTAGTGT TATTTCGAGT AGGTTTTATG GGTGAACAAC CCTTGACCGC  360
CAAATCAATC ACAAGAGTTC AACATTTTAT TTATTTTGAA ATGTATTAAA AATCGTTAAT  420
CTATATATTC GCCCCATTAT TGGGATTAAA TATTCACAAG GGTTTAGACC GTCATGAGAC  480
AGATTAGTTT TATCTTACTG ATGGTCACAT CACAATAGTA ATTCAACTTA ATACGAGAGG  540
AACCATTGAT TCACGCAATT GGTCATCGCA CTTAGTTGAA AAGCTAGGGG TGCGAAGCTA  600
CCGTACGCTG GATTATGATT GAACACCTCT AAGTCAGAAT CCGAATTAGA AACAATGCAC  660
GTGTCCGTTG CCTGATTGCC AACCCCAATA ACACGTGTTG TAGGTTTAAC CATGTTTATG  720
AAAGATAAGG TTTTTTTTTT TATAAGCAAG CAACTATAGG GGTTTACTTC CGTGCGCAAA  780
TTTTTAGGTT ACCTATTTTG GGAGGGGGGA TTATGATTCA AGTGAAAGAA AGTTGGCACA  840
CACACAATCA GTACATCTGT TTTGACAGAG ACACAGCCTA AAAACAGCAG CAAACAAGCC  900
TAAAGGAATC ACCCAAAAAC AACAACCAAA AGTACAGAGG AAAACAAAAG AATCCCTGTT  960
ACCACCAAGC TGAAAAAAAG AAAATAAAAC TCAACTTTTG GCAATAAAAA CCCTCCTACC 1020
CTCAACCCCT AACCACGCAA CAATCAGCAA TACTCCAAGC AACCATTTTC CTTACAAGTT 1080
```

FIGURE 7A

```
TGTTTTTCTT GTGATTAATC CAT ATG GCT AGC TCC ATG TCC CTT AAG CTT GCA 1133
                          Met Ala Ser Ser Met Ser Leu Lys Leu Ala>

TGT CTG CTA GTG TTG TGC ATG GTG GTG GGT GCA CCC CTG GCT CAA GGG   1181
Cys Leu Leu Val Leu Cys Met Val Val Gly Ala Pro Leu Ala Gln Gly>

GAC GTA ACC CGT GCT GAT GGC GTA GTC ACC CTT CCA CGC TGC CTT CCT   1229
Asp Val Thr Arg Ala Asp Gly Val Val Thr Leu Pro Arg Cys Leu Pro>

TTA TTG ATA GGG AAT GGT AAT GGT GCT GAT GCT GAT GTT GAT GCC CCA   1277
Leu Leu Ile Gly Asn Gly Asn Gly Ala Asp Ala Asp Val Asp Ala Pro>

GCT TGC TGC GAC ATC GTC AGG GGT CTC TTG AGC TCG CTG CTC TGT GGT   1325
Ala Cys Cys Asp Ile Val Arg Gly Leu Leu Ser Ser Leu Leu Cys Gly>

GGT GTT TAGGAACCG ATCTAGCTTG AAATCGGGTT CGGATACGGG TGGAGTTTCA     1380
Gly Val>

AATTGGTGTG TTATGGAATC CCAACTTAAT CGTGTTTAGG GGTGGGATCC AATTGTGTGA 1440

TACATTACAG AGCATGGTTG TGGATTGTTT TCTCATATGT TTTGATTGAC TTGCTTGATA 1500

CATTGGATGA TTCGATAAGG TGACCGGTTT ACCTGGGTAT CCAACCATCA TCCGATTACT 1560

TTTTAATAAT TATTTGTTTC TTCTTTATGT TGTCTGTCTT TTTGTTTCTT GATCTATAAC 1620

ATTATATTTG CCCAAATTTT CGCATTTTCC ATATGTAGCT TATATATGTA TATATATATT 1680

CAATAAAGTA TATTGATTTA GCAGATGATT TGTGTATATA TTTAAATCAA ATCAAACATT 1740

AATGATCATT CACTAGCGTC TTAATCTTGA AAAATTCATC AACGGTTATC CTTTGCAGCA 1800

TATATAAAAA AAATTGCCAA CCCTATGCTT TTACACCTAA TTCAAGGGAT AACATAAGTC 1860

GATTAAAACG A                                                      1871
```

FIGURE 7B

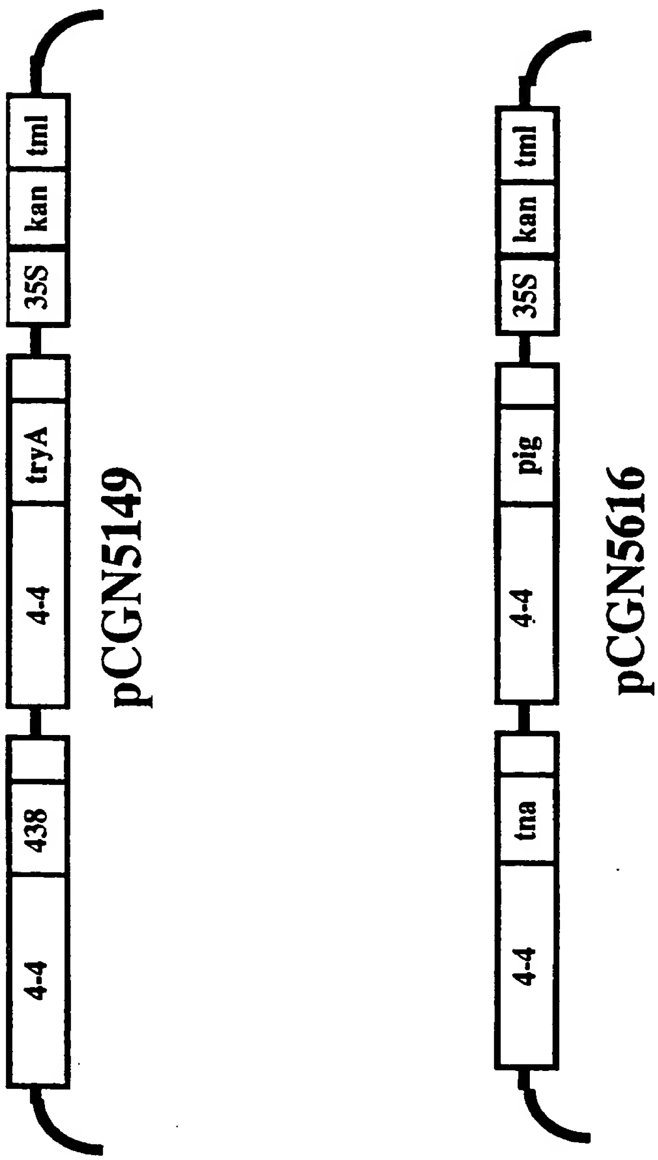
tml
kan
35S
tryA
4-4
438
4-4
pCGN5149
tml
kan
35S
pig
4-4
tna
4-4
pCGN5616

FIGURE 8

Coker 130	Yxy, Y	Yxy, x	Yxy, y	Lab, L	Lab,a	Lab,b	LCh, L	LCh,C	LCh, h
1	80.35	.3206	0.3266	91.84	0.16	5.51	91.84	5.51	88.4
2	77.62	.3232	0.3282	90.6	0.66	6.45	90.6	6.48	84.2
3	80.98	.3197	0.3257	92.12	0.13	5.04	92.12	5.04	88.6
4	80.16	.3200	0.3255	91.75	0.35	5.00	91.75	5.01	86.1
5	77.03	.3220	0.3271	90.33	0.61	5.84	90.33	5.87	84.1
6	73.67	.3258	0.3293	88.76	1.35	7.14	88.76	7.26	79.4
7	82.43	.3178	0.3237	92.76	0.15	4.05	92.76	4.05	87.9
8	82.21	.3196	0.3255	92.66	0.19	4.99	92.66	4.99	87.9
9	81.19	.3194	0.3241	92.21	0.77	4.42	92.21	4.48	80.2
10	76.11	.3243	0.329	89.9	0.74	6.89	89.9	6.92	84
11	82.28	.3178	0.3236	92.69	0.19	4.00	92.69	4.00	87.3
TOTAL	874.03	3.5302	3.5883	1005.62	5.30	59.33	1005.62	59.61	938.10
MEAN	79.46	.3209	.3262	91.42	0.48	5.39	91.42	5.42	85.28
S.D.	2.91	.0026	.0020	1.33	0.38	1.08	1.33	1.11	3.22
RANGE	82.43-73.67	.3858-.3178	0.3293-.3236	92.76-88.76	1.35-.13	7.14-4.00	92.76-88.76	7.26-4.00	88.6-79.4
AVER DEV.	2.44	.0021	.0017	1.11	0.31	0.88	1.11	0.90	2.64

Coker 130	Hunter L	Hunter a	Hunter B
1	89.63	0.15	5.42
2	88.10	0.66	6.27
3	89.98	0.13	4.98
4	89.53	0.36	4.94
5	87.76	0.61	5.69
6	85.83	1.35	6.85
7	90.79	0.15	4.03
8	90.67	0.19	4.95
9	90.10	0.78	4.38
10	87.23	0.75	6.65
11	90.70	0.19	3.98
TOTAL	980.32	5.32	58.14
MEAN	89.12	0.48	5.29
S.D.	1.65	0.39	0.99
RANGE	90.79-85.83	1.35-.13	6.85-3.98
AVER DEV.	1.37	0.31	0.81

FIGURE 9

5148	Yxy, Y	Yxy, x	Yxy, y	Lab, L	Lab,a	Lab,b	LCh, L	LCh,C	LCh, h
68-1	60.76	0.34	0.35	82.24	2.32	15.11	82.24	15.28	81.3
68-1	61.89	0.34	0.34	82.82	1.97	14.31	82.85	14.44	82.2
50-2-1	78.39	0.3324	0.3375	90.95	0.68	11.29	90.95	11.31	86.6
50-2-1									
(lint fiber)	21.49	.3155	0.3489	53.48	-8.01	7.97	53.48	11.29	135.2

5148	Hunter L	Hunter a	Hunter B
68-1	77.94	2.25	13.35
68-1	78.67	1.92	12.75
50-2-1	88.53	0.69	10.71
50-2-1			
(lint fiber)	46.35	-6.35	6.06

FIGURE 10

5149	Yxy, Y	Yxy, x	Yxy, y	Lab, L	Lab,a	Lab,b	LCh, L	LCh,C	LCh, h
68-1	65.75	0.3351	0.34	84.86	0.72	11.9	84.86	11.92	86.6
68-1	62.54	.3458	0.3474	83.19	2.14	15.84	83.19	15.98	82.4
68-1	62.56	0.3458	0.3474	83.2	2.14	15.85	83.2	15.99	82.4
8-1	84.72	.3196	0.3278	93.76	0.89	5.87	93.76	5.93	98.6
68-1	64.97	.3316	0.3354	84.46	1.17	9.81	84.46	9.87	83.3
17-2	64.42	.3423	0.3436	84.18	2.26	14.19	84.18	14.36	81
17-3	60.97	.3475	0.3475	82.36	2.74	16.03	82.36	16.26	80.4
17-15-1	64.02	.3433	0.3444	83.97	2.34	14.57	83.97	14.75	80.9
21-1	59.32	0.3443	0.3445	81.46	2.64	14.41	81.46	14.64	79.7
21-3	63.64	0.34	0.3409	83.77	2.4	12.89	83.77	13.11	79.5
21-6	67.12	0.3372	0.3394	85.56	1.88	12.15	85.56	12.29	81.3
50-3-1	61.26	0.3502	0.3511	82.51	2.4	17.63	82.51	17.79	82.3
67-1	64.34	0.3434	0.3442	84.13	2.48	14.58	84.13	14.78	80.4
68-1	64.12	0.3442	0.3447	84.02	2.58	14.85	84.02	15.07	80.2
68-2	70.21	0.3428	0.3447	87.09	2.05	15.04	87.09	15.17	82.3
68-3	63.81	0.3457	0.3468	83.86	2.35	15.76	83.86	15.93	81.6

5149	Hunter L	Hunter a	Hunter B
68-1	81.08	0.71	10.89
68-1	79.08	2.08	14
68-1	79.09	2.09	14.02
8-1	92.04	0.91	5.81
68-1	80.6	1.15	9.06
17-2	80.25	2.21	12.75
17-3	78.08	2.68	14.09
17-15-1	80.01	2.29	13.05
21-1	77.01	2.56	12.73
21-3	79.77	2.35	11.65
21-6	81.92	1.86	11.14
50-3-1	78.26	2.33	15.36
67-1	80.2	2.43	13.07
68-1	80.07	2.53	13.28
68-2	83.79	2.04	13.68
68-3	79.87	2.3	14

FIGURE 11

5616	Yxy, Y	Yxy, x	Yxy, y	Lab, L	Lab,a	Lab,b	LCh, L	LCh,C	LCh, h
11-1	72.26	0.3215	0.3254	88.09	1.1	5.06	88.09	5.17	77.8
11-2	58.69	0.3284	0.3335	81.12	0.6	8.36	81.12	8.38	85.9
11-2	52.78	0.3358	0.3335	77.74	3.55	9.22	77.74	9.87	69
11-1	72.03	0.3312	0.3338	87.98	1.72	9.52	87.98	9.67	79.8
11-1	72.34	0.3295	0.332	88.13	1.79	8.64	88.13	8.82	78.4
11-1	71.98	0.3295	0.3313	87.95	2.09	8.39	87.95	8.64	76.1
11-1	73.01	0.3256	0.3305	88.45	0.68	7.51	88.45	7.54	84.9
17-1-2	75.85	0.3274	0.3306	89.78	1.52	7.94	89.78	8.08	79.3
17-3-1	72.6	0.3271	0.3303	88.25	1.48	7.66	88.25	7.8	79.1
17-4-1	69.02	0.3352	0.3377	86.51	1.78	11.37	86.51	11.5	81.2
25-11-1	69.5	0.3364	0.3401	86.75	1.26	12.41	86.75	12.47	84.2
25-28-1	72.21	0.3324	0.3343	88.06	2.09	9.9	88.06	10.11	78.2
25-36-2	70.46	0.3327	0.3353	87.22	1.73	10.22	87.22	10.36	80.5
35-35-1	75.59	0.3268	0.3299	89.66	1.56	7.58	89.66	7.73	78.4
50-12-1	73.13	0.3284	0.3316	88.5	1.46	8.36	88.5	8.48	80.1
KS-11-2	65.33	0.3371	0.3388	84.65	2.07	11.83	84.65	12	80.1

5616	Hunter L	Hunter a	Hunter B
11-1	85	1.09	4.89
11-2	76.61	0.58	7.64
11-2	72.64	3.38	8.22
11-1	84.87	1.72	8.97
11-1	85.05	1.79	8.2
11-1	84.84	2.08	7.96
11-1	85.44	0.67	7.18
17-1-2	87.08	1.52	7.62
17-3-1	85.2	1.48	7.31
17-4-1	83.07	1.76	10.52
25-11-1	83.36	1.25	11.43
25-28-1	84.97	2.08	9.32
25-36-2	83.94	1.72	9.56
35-35-1	86.94	1.57	7.29
50-12-1	85.51	1.46	7.96
KS-11-2	80.82	2.04	10.81

FIGURE 12

BC	Yxy, Y	Yxy, x	Yxy, y	Lab, L	Lab,a	Lab,b	LCh, L	LCh,C	LCh, h
12 Gr en	33.34	0.3779	0.3717	66.01	4.24	24.18	66.01	24.54	80.1
22 Brown	38.18	0.3778	0.3662	68.15	6.18	23.31	68.15	24.11	75.2
3 Red	24.23	0.4055	0.3728	56.31	10.96	25.52	56.31	27.77	66.9
4 Ivory	46.84	0.3657	0.3599	74.08	4.6	21.13	74.08	21.62	77.8

BC	Hunter L	Hunter a	Hunter B
12 Gre n	59.44	3.79	17.92
22 Brown	61.78	5.62	17.69
3 Red	49.22	9.42	17.14
4 Ivory	68.43	4.31	17.02

FIGURE 13

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/29, 15/82, 5/10, A01H 5/00	A3	(11) International Publication Number: WO 96/40924 (43) International Publication Date: 19 December 1996 (19.12.96)

(21) International Application Number: PCT/US96/09897

(22) International Filing Date: 7 June 1996 (07.06.96)

(30) Priority Data:
08/480,178 7 June 1995 (07.06.95) US

(60) Parent Application or Grant
(63) Related by Continuation
US Not furnished (CIP)
Filed on Not furnished

(71) Applicant *(for all designated States except US)*: CALGENE, INC. [US/US]; 1920 Fifth Street, Davis, CA 95616 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: McBRIDE, Kevin [US/US]; 1309 Marina Circle, Davis, CA 95616 (US). STALKER, David, M. [US/US]; 2736 Cumberland Place, Davis, CA 95616 (US). PEAR, Julie, R. [US/US]; 818 Douglass Avenue, Davis, CA 95616 (US). PEREZ-GRAU, Luis [ES/US]; 1230 Elk Place, Davis, CA 95616 (US).

(74) Agents: SCHWEDLER, Carl, J. et al.; Calgene, Inc., 1920 Fifth Street, Davis, CA 95616 (US).

(81) Designated States: AU, CA, CN, JP, KG, KZ, MX, TJ, TM, TR, US, UZ, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report: 6 February 1997 (06.02.97)

(54) Title: COTTON FIBER TRANSCRIPTIONAL FACTORS

(57) Abstract

Novel DNA constructs are provided which may be used as molecular probes or inserted into a plant host to provide for modification of transcription of a DNA sequence of interest during various stages of cotton fiber development. The DNA constructs comprise a cotton fiber transcriptional initiation regulatory region associated with a gene which is expressed in cotton fiber. Also provided is novel cotton having a cotton fiber which has a natural color introduced by the expression in the cotton fiber cell, using such a construct, of pigment synthesis genes. Cotton fiber cells having color produced by genetic engineering and cotton cells comprising melanin and indigo pigments are included.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

INTERNATIONAL SEARCH REPORT

Int ional Application No
PCT/US 96/09897

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/29 C12N15/82 C12N5/10 A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO,A,95 08914 (AGRACETUS) 6 April 1995 see page 8, line 12 - page 9, line 17 see page 18, line 6 - page 20, line 16 see page 36, line 1 - page 38, line 36 see sequence ID nos 4 and 5 ---	1
X	WO,A,94 12014 (AGRACETUS) 9 June 1994 see page 9, line 29 - page 10, line 32 see page 19, line 5 - page 21, line 24 see page 40, line 8 - page 43, line 35 --- -/--	1

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

2

Date of the actual completion of the international search	Date of mailing of the international search report
6 December 1996	20.12.96

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Maddox, A

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Int. ...onal Application No
PCT/US 96/09897

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE WPI Section Ch, Week 9021 Derwent Publications Ltd., London, GB; Class A35, AN 90-161489 XP002020420 & JP,A,02 104 773 (NAGASE SANGYO KK) , 17 April 1990	38
A	see abstract	13-21, 27-36
X	TRENDS IN BIOTECHNOLOGY, vol. 10, no. 5, 1 May 1992, pages 165-170, XP000272383 JOHN M E ET AL: "GENES FOR JEANS: BIOTECHNOLOGICAL ADVANCES IN COTTON" see figure 2C	38
P,X	WO,A,95 16783 (CALGENE INC ;MCBRIDE KEVIN E (US); STALKER DAVID M (US)) 22 June 1995 see page 8, line 38 - page 10, line 14	13-18,20
P,X	US,A,5 495 070 (JOHN MALIYAKAL) 27 February 1996 see sequence ID no. 4. see column 27 - column 28; example 3 see column 45 - column 46	23,26
P,X	MOLECULAR AND GENERAL GENETICS, vol. 248, July 1995, pages 43-51, XP002018965 DELMER, D.P., ET AL.: "Genes encoding small GTP-binding proteins analagous to mammalian rac are preferentially expressed in developing cotton fibers" see the whole document	24
A	EP,A,0 409 629 (CALGENE INC) 23 January 1991 see the whole document	1
A	CROP SCI 25 (5). 1985. 792-797. , XP000612251 KOHEL R J: "GENETIC ANALYSIS OF FIBER COLOR VARIANTS IN COTTON GOSSYPIUM-HIRSUTUM." see the whole document	13-21, 27-36
A	TWELFTH ANNUAL CONFERENCE OF THE SOCIETY OF ETHNOBIOLOGY, RIVERSIDE, CALIFORNIA, USA, MARCH 30-APRIL 2, 1989. J ETHNOBIOL 9 (2). 1989. 239. , XP000612252 FOX S V: "NATURALLY OCCURRING COTTON LINT COLORS." see the whole document	13-21, 27-36

-/--

2

Int ional Application No
PCT/US 96/09897

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CHEMISTRY AND INDUSTRY, 5 September 1994, pages 676-679, XP000466067 JOHN, M.E.: "Re-engineering cotton fibre" see page 677, left-hand column, paragraph 2; table 2	13-21, 27-36
A	WO,A,89 08145 (BIOSOURCE GENETICS CORP) 8 September 1989 see page 5, line 8 - line 18 see page 28, line 17 - line 20 see page 41, line 21 - page 42, line 2 see page 49, line 8 - line 17	16,17
A	WO,A,95 13386 (BIOSOURCE TECH INC) 18 May 1995 see page 23, line 26 - page 24, line 35	16,17
A	WO,A,93 02195 (COMMW SCIENT IND RES ORG) 4 February 1993 see page 1, line 23 - line 29 see page 1, line 36 - page 2, line 2 see page 7, line 1 - line 7	16,17
A	EP,A,0 198 288 (ADVANCED GENETIC SCIENCES INC) 22 October 1986 see page 7, line 32 - line 55	18
T	US,A,5 530 185 (MARTINEAU BELINDA M ET AL) 25 June 1996 see column 8, line 23 - column 9, line 14; claims 1-18; example 10	13,14, 16-20
T	WO,A,96 26639 (CALGENE INC) 6 September 1996 see page 20, line 8 - page 21, line 17 see page 47, line 11 - page 49, line 6	13,14, 16,17

2

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int ional Application No
PCT/US 96/09897

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A-9508914	06-04-95	AU-A-	7925094	18-04-95
		EP-A-	0670670	13-09-95
		JP-T-	8509871	22-10-96
WO-A-9412014	09-06-94	AU-A-	5676394	22-06-94
		CA-A-	2127807	09-06-94
		EP-A-	0625006	23-11-94
WO-A-9516783	22-06-95	US-A-	5576198	19-11-96
		CA-A-	2178645	22-06-95
		EP-A-	0734446	02-10-96
US-A-5495070	27-02-96	US-A-	5521078	28-05-96
EP-A-0409629	23-01-91	WO-A-	9101324	07-02-91
		US-A-	5175095	29-12-92
		US-A-	5530185	25-06-96
WO-A-8908145	08-09-89	AU-B-	638411	01-07-93
		AU-A-	4072589	22-09-89
		EP-A-	0406267	09-01-91
		JP-T-	3502886	04-07-91
		US-A-	5529909	25-06-96
		US-A-	5316931	31-05-94
WO-A-9513386	18-05-95	US-A-	5529909	25-06-96
		AU-A-	1173495	29-05-95
		EP-A-	0731843	18-09-96
		ZA-A-	9408973	08-02-96
WO-A-9302195	04-02-93	AU-B-	655494	22-12-94
		CA-A-	2112998	04-02-93
		EP-A-	0599868	08-06-94
		JP-T-	7501686	23-02-95
		NZ-A-	243594	26-08-94
EP-A-0198288	22-10-86	US-A-	5180873	19-01-93
		AU-B-	602306	11-10-90
		AU-A-	5615586	23-10-86
		US-A-	5278057	11-01-94

Form PCT/ISA/210 (patent family annex) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 96/09897

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US-A-5530185	25-06-96	US-A- 5175095	29-12-92
		EP-A- 0409629	23-01-91
		WO-A- 9101324	07-02-91
WO-A-9626639	06-09-96	AU-A- 5363296	18-09-96

Form PCT/ISA/210 (patent family annex) (July 1992)